Constraining the parameters of GW150914 & GW170104 with numerical relativity surrogates

Prayush Kumar,[1,*] Jonathan Blackman,[2] Scott E. Field,[3] Mark Scheel,[2] Chad R. Galley,[4,2] Michael Boyle,[1] Lawrence E. Kidder,[1] Harald P. Pfeiffer,[5] Bela Szilagyi,[2,4] and Saul A. Teukolsky[1,2]

[1] *Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*
[2] *Theoretical Astrophysics, Walter Burke Institute for Theoretical Physics, MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA*
[3] *Department of Mathematics, University of Massachusetts, Dartmouth, MA 02747, USA*
[4] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA*
[5] *Albert Einstein Institute, Am Mühlenberg, Golm, Germany*

(Dated: August 27, 2018)

Gravitational-wave detectors have begun to observe coalescences of heavy black holes at a consistent pace for the past few years. Accurate models of gravitational waveforms are essential for unbiased and precise estimation of the source parameters, such as masses and spins of component black holes. Recently developed surrogate models based on high-accuracy numerical relativity simulations provide ideal models for constraining physical parameters describing these heavy black hole merger events. In this paper we demonstrate the viability of these surrogate models as reliable parameter estimation tools, and show that within a fully Bayesian framework surrogates can help us extract more information from gravitational wave observations than traditional models. We demonstrate this by analyzing a set of synthetic signals and showing the improvement that the use of numerical relativity surrogates bring to our parameter estimates. We then consider the case of two of the earliest binary black holes detected by the LIGO observatories, GW150914 and GW170104, and reanalyze their data with a generically precessing numerical-relativity-based surrogate model. For these systems we find that overall results are quantitatively consistent with inferences performed with conventional models, except that our refined analysis estimates the sources of both GW150914 and GW170104 to be $10-20\%$ further away than previously estimated and constrain their orientation to be closer to either face-on or face-off configurations more strongly than in the past. Additionally, for GW150914 we constrain the effective spin parameter to be closer to zero. This work is a first step toward eliminating the approximations used in semi-analytic waveform models from GW parameter estimation.

PACS numbers: 04.30.-w, 04.30.Db

I. INTRODUCTION

General Relativity (GR) predicts that accelerated massive bodies emit energy in the form of gravitational waves (GWs). In 2015, the first direct detection of GWs coming from coalescing binary black holes (BBHs) was made by the LIGO observatories [1]. Since then, many more GW signals from BBHs have been observed by the LIGO and Virgo detectors [2–5], ushering us into the era of GW astronomy. GW searches for BBH signals [6–8], the process of estimating their source properties [9], as well as that of testing GR with them [10] rely heavily on the technique of matched filtering, which tacitly assumes the availability of GW signal models for BBHs.

For heavy black holes, such as those that have dominated the event rates of LIGO-Virgo detectors so far [3], a large fraction of the observable signal consists of the last few tens of orbits prior to the binary's merger. In this regime, the dynamical effects of GR are substantial, making analytic treatment difficult. Instead, numerical solutions of Einstein's equations [11–16] must be used.

The inspiral of a binary system of black holes along a quasi-circular trajectory, and their subsequent merger and ringdown, is completely describable by 8 parameters - the masses of both holes, and their spin vectors. Conventional parameter estimation (PE) algorithms search through this parameter space to estimate parameters that best describe the signal embedded in LIGO-Virgo data. In the process, they can take a large number ($\mathcal{O}(10^6)$) of steps, each requiring evaluation of a new waveform. Although we have the technical capability to perform full numerical relativity (NR) simulations over a good fraction of the multi-dimensional parameter space, each simulation still takes a large amount of computing and human time. Therefore, it has remained impractical to use NR simulations directly with conventional PE methods for estimating physical parameters of BBHs. There are two possible alternatives that have both been demonstrated in the past: (a) using phenomenological waveform models containing free parameters that are tuned to a (relatively) small number of NR simulations [17, 18]; and (b) using grid-based parameter estimation methods [19]. Phenomenological waveform models (alternative (a)) have been used in previous LIGO-Virgo publications, but have also been shown to break down close to the parameter-space boundary of their calibra-

* prayush@astro.cornell.edu

tion domain [20, 21] or even within it [22]. This requires a recalibration of the model. Grid-based PE methods (alternative (b)) have been recently applied to GW observations [23, 24]; however, they have so far only demonstrated the ability to constrain a subset of the 7 physical parameter because of the sparse parameter space coverage of available NR simulations.

A novel alternative arises from the development of surrogate models [25, 26] for numerical relativity waveforms [27]. Such data-driven models are constructed over a training set of specially selected NR simulations. The waveforms from these simulations are then "interpolated" in parameter space. The resulting NR surrogate model is able to quickly generate new waveforms at arbitrary points within the training region with the largest surrogate model errors typically comparable to the largest errors in the numerical relativity simulations. After a couple of simpler versions [27, 28], Blackman et al [29] published a surrogate model `NRSur7dq2` for generically spinning-precessing binaries. This `NRSur7dq2` model was developed using 744 new simulations [27] spanning a range of the 7-dimensional space[1] bounded in mass ratio $q \leq 2$, and BH spin magnitudes $|\vec{\chi}_{1,2}| \leq 0.8$. In this paper, we demonstrate the viability of using this `NRSur7dq2` model in the follow-up parameter estimation of GW signals coming from heavy BBHs such as those that LIGO-Virgo observe regularly. In particular, we use it to estimate all physical parameters of the first two heavy BBH events GW150914 and GW170104, significantly extending their past analyses [4, 23, 30].

We first perform controlled tests by injecting 48 synthetic GW signals into zero noise and inferring the signal's parameter values with the `NRSur7dq2` surrogate model. We compare with the results of the IMRPhenomPv2 model, which captures spin-orbit precession effects and has been used in previous LIGO-Virgo analyses [31–33]. We vary the source parameters as follows. Mass ratio takes on distinct values $q \in \{1.2, 1.5\}$, source location is varied from $500 - 1500$Mpc, inclination is varied between close to face-on and edge-on, and component spins are chosen from four distinct configurations with magnitudes $0.4 - 0.65$. The values are deliberately chosen to enhance spin-induced orbital precession. We find that the parameter recovery of `NRSur7dq2`, both with and without sub-dominant $\ell > 2$ waveform modes, is broadly consistent with IMRPhenomPv2. In particular, sky location, binary chirp mass and its precessing-spin components are recovered remarkably similarly by all models. We also find that the inclusion of higher order $\ell = \{3, 4\}$ GW modes enhances our ability to measure the luminosity distance to the source and its orbital inclination, its total mass as well as its effective spin. These improvements are incremental and we expect them to be greater for binary sources with higher mass ratios for which sub-dominant modes carry a larger fraction of the total signal power [34–36]. Since `NRSur7dq2` is restricted to $1 \leq q \leq 2$, these results provide incentive for extending its range of validity to higher mass ratios.

Having established the viability of `NRSur7dq2` surrogate within a fully Bayesian parameter recovery framework, we next analyze the first-ever recorded BBH merger event: GW150914. The primary improvement we note is in the estimation of the binary's luminosity distance d_L from Earth: with extra information coming from sub-dominant modes, we are able to constrain d_L close to $\sim$ 530Mpc, about 100Mpc further away than all others models' estimation. Simultaneously, the surrogate also constrains the source of GW150914 to be either face-on or face-off more strongly than other models, disfavoring edge-on configurations. Consistent with this, the `NRSur7dq2` models estimate the source's chirp/total mass to be marginally higher than what approximate models measure. And finally, having complete 2-body spin information encoded in them, the `NRSur7dq2` models constrain the effective-spin of GW150914 to be closer to zero than other models (with the same sampling priors). These results continue to hold when we compare them with the LVC analysis of this event [37].

We also analyze the second heavy BBH event GW170104 recorded by LIGO detectors in early 2017. As for GW150914, we find that the surrogate constrains the luminosity distance to this event to be larger (by 10%) than what approximate models that include only $\ell = |m| = 2$ GW modes do. Similarly, it also constrains the source to be closer to face-on/off than edge-on more strongly than other models. The estimation of mass parameters is consistent between `NRSur7dq2` and semi-analytic models, with the former only recovering the portion of mass ratio posterior with support in $1 \leq q \leq 2$. Finally, we find the estimation of spin parameters to be remarkably similar between `NRSur7dq2`, IMRPhenomPv2 and SEOBNRv4, with little extra information coming from the use of NR surrogate. These results are all consistent with the the first analysis of this event by the LVC [4]. A summary in the form of median estimates and symmetric 90% credible intervals for inferred quantities is given in Table II.

From the analyses of GW150914 and GW1701014, we learn that one consistently recovers additional information that helps break the luminosity distance - inclination degeneracy for BBH events, allowing us to constrain GW source location better. We also learn that, in some cases, one could constrain BH spins better with the NR surrogates since they contain unabbreviated spin degrees of freedom. However, spin measurements can be sensitive to the choice of sampling priors employed [30, 38], and we defer an investigation of their effect on spin inferences for both events to future work. Our results are encouraging and we propose that `NRSur7dq2` and future NR surrogate models be used as part of standard GW event follow-ups. In order to enable further analysis by the community, we provide full posterior samples from

[1] All simulations can be re-scaled to any point in the eight dimension of total mass M.

Bayesian parameter estimation of LIGO/Virgo data for GW150914 and GW170104, with `NRSur7dq2HM` and IMRPhenomPv2.

The remainder of this paper is organized as follows. In Section II we describe the surrogate and approximate waveform models used in this paper, as well as the details of our Bayesian parameter estimation machinery. In Section III we present results from studies involving parameter recovery from synthetic signals. In Section IV and V we present results of our re-analysis of GW150914 and GW1701014 using the new NR surrogate model. And finally, in Section VI we summarize our findings and present the future outlook for this research.

II. PRELIMINARIES

A. Numerical Relativity Surrogates

A surrogate waveform model is one that takes a set of pre-computed waveforms generated by an underlying model as input, and interpolates in parameter space between these waveforms to produce waveforms for arbitrary parameter values. The underlying waveform model can be analytic, phenomenological, or purely numerical. Surrogates can often be evaluated in a fraction of the time that takes for the underlying model to generate a waveform, and was in fact originally proposed as a way to reduce the computational cost of otherwise expensive waveform models when used with MCMC-based parameter estimation algorithms applied to GW events [25]. With interpolation comes an additional source of modeling error, called the *surrogate* error. In principle, this error can be arbitrarily reduced by using a sufficiently large set of pre-computed waveforms to cover the parameter space. In practice, when using NR waveforms the cost may become prohibitive.

The `NRSur7dq2` model of Ref. [29] spans the 7-dimensional space of spin-precessing non-eccentric black hole binaries. It is built from the results of 744 NR simulations performed using the Spectral Einstein Code `SpEC` [39] and has already found several applications [40, 41]. That it spans all spin-precession degrees of freedom comes at the cost of limiting its domain to comparable mass ratios $q = m_1/m_2 \leq 2$ and black hole spins with magnitudes ≤ 0.8 of their extremal values. The choice of NR simulations used to train this surrogate was based on a combination of methods including sparse grids [42, 43] (as detailed in Appendix A of [29]), a template-metric-type stochastic sampler, and existing NR simulations. Taken together, these choices maximized the coverage of the binary parameter space with as few simulations as possible while simultaneously keeping the surrogate error sufficiently small. Instead of modeling waveform modes directly across the parameter space, the strategy of `NRSur7dq2` is to interpolate quantities that have as little structure (such as oscillations) as possible. Ref. [29] constructs surrogate models for combinations of waveform modes in the *coorbital frame*, as well as for orbital phase and spin-related quantities that are required to transform these modes back to an inertial frame. They choose to parameterize these fits using *instantaneous* spins and mass ratio, instead of *initial* spins, as they find this choice improves the quality of fits. Therefore, evaluation of `NRSur7dq2` requires first obtaining the full time-evolution of BH spins, orbital phase, and the unit quaternion that defines the coprecessing frame [31, 44, 45]; and subsequently using these evolutions to construct full inertial frame waveform modes from surrogate evaluation of coorbital frame modes. We refer the reader to [28, 29] for further technical details and reasoning supporting various choices of surrogate construction. Finally, we note that `NRSur7dq2` is limited in length to span the last 20 binary orbits before merger. In practice, with a lower frequency cutoff of 20 Hz, this restricts its use to binaries with total masses $M = m_1 + m_2 \gtrsim 50 M_\odot$. In the remainder of the paper, we will use `NRSur7dq2` to mean the surrogate with only $\ell = 2$ waveform modes included, `NRSur7dq2HM` for the surrogate with all its available $\ell = \{2, 3, 4\}$ modes, and `NRSur7dq2` when discussing the surrogate model in general. [This appears a bunch, for example, at the beginning of section 3]

B. Analytic Waveform Models

In this paper, we will consider two waveform families: Effective-One-Body (EOB) and phenomenological (IMRPhenom) [18, 46]. Both of these are semi-analytic models of the complete inspiral-merger-ringdown for spinning BHs with non-eccentric orbits. For both models, we consider the dominant $(\ell, m) = (2, \pm 2)$ spin-weighted spherical harmonic waveform multipoles as only these have been calibrated to NR simulations through merger.

Effective-One-Body: The effective-one-body approach solves for the dynamics of the two-body problem in nonlinear GR by mapping it to the dynamics of an *effective* test particle of mass $\mu = m_1 m_2/(m_1 + m_2)$ and spin $S^*(m_1, m_2, \vec{\chi}_1, \vec{\chi}_2)$ in a background spacetime that is described by a parameterized deformation of the Kerr metric. Both S^* and the background deformation (to leading order) are chosen so that the geodesic followed by the test particle reproduces the perturbative post-Newtonian (PN) dynamics of the original two-body system [47]. This conserved dynamics of the test particle is described by the EOB Hamiltonian, which is also derived to leading order using PN results. The radiative dynamics is introduced through a flux of energy to emitted gravitational radiation, obtained by summing over all PN-expanded waveform modes at future null infinity. All of these model pieces are individually taken beyond known PN orders through resummation and addition of phenomenological parameters that are subsequently calibrated to ensure agreement of the inertial-frame waveform multipoles with NR simulations. This allows the EOB prescription to be extended beyond the slow-motion

TABLE I. Parameters of injection set. Total $1 \times 2 \times 4 \times 2 \times 3 = 48$ injections. For each injection, the polarization angle is uniformly sampled on the interval $[0, \pi]$ and the right ascension and declination are uniformly sampled on the two-sphere. Their combined network SNR ranges from $\rho = 13 - 83$ for the two-detector LIGO network.

$M \equiv m_1 + m_2\,(M_\odot)$	$q \equiv m_1/m_2$	$(\vec{\chi}_1, \vec{\chi}_2)$	$\theta_{\rm JN}$	D (Mpc)	ψ	(α, δ)	Signal model
60	$\{1.2, 1.5\}$	$\left\{ \left(\frac{0.65}{\sqrt{2}}(1,1,0), -\frac{0.65}{\sqrt{2}}(1,1,0)\right), \left(\frac{0.65}{\sqrt{2}}(1,1,0), (0,0,0)\right), \left(\frac{0.4}{\sqrt{3}}(1,1,1), \frac{0.4}{\sqrt{3}}(1,1,1)\right), \left(\frac{0.4}{\sqrt{3}}(1,1,1), -\frac{0.4}{\sqrt{3}}(1,1,1)\right) \right\}$	$\left\{\frac{\pi}{6}, \frac{5\pi}{12}\right\}$	$\{500, 1000, 1500\}$	$U(0,\pi)$	$U(S^2)$	`NRSur7dq2HM`

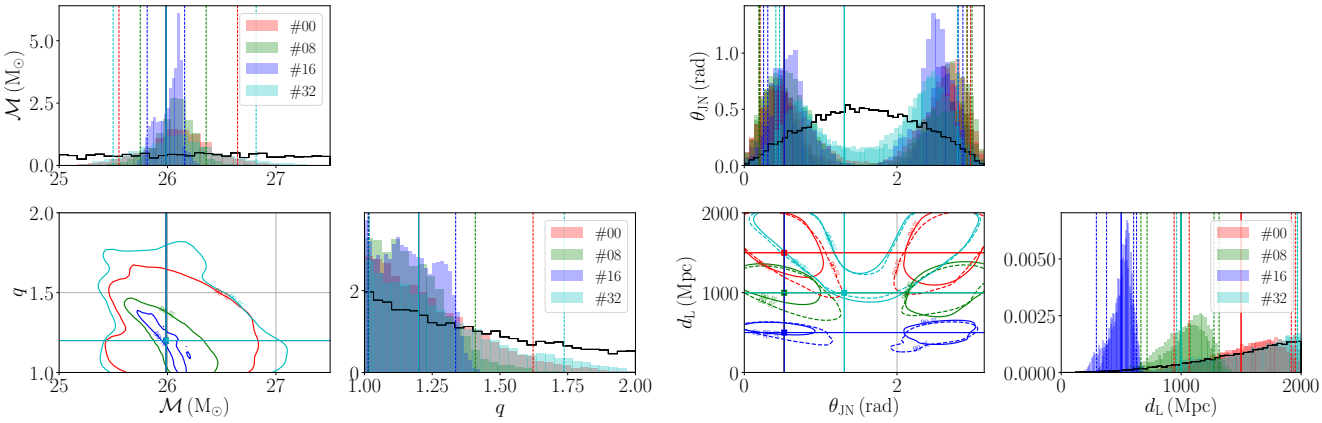

FIG. 1. Parameter recovery for select synthetic injections with `NRSur7dq2HM`. Here we show the mass (left) and source distance / inclination angle (right) recovery for four injections, which are labeled $\#0, 8, 16, 32$ in Fig. 2. All 4 injections $(\#0, 8, 16, 32)$ have identical source mass and spin parameters: $M = 60M_\odot$, $q = 1.2$, $\vec{\chi}_1 = -\vec{\chi}_2 = \frac{0.65}{\sqrt{2}}(1, 1, 0)$. The first three have identical source inclination angles $\theta_{\rm JN} = \pi/6$, and differ in the distance at which their source is located, which is 1500Mpc, 1000Mpc, and 500Mpc respectively. The fourth synthetic signal (labeled $\#32$) is similar to signal $\#8$ with the only difference being that the binary's inclination angle with respect to line of sight is now much closer to edge-on, i.e. $\theta_{\rm JN} = 5\pi/12$ for $\#32$. Only in the right panel, we additionally show results from the analysis of the corresponding synthetic signal with `NRSur7dq2` templates (without $l > 2$ modes) as dashed 2-D contours, which can be directly contrasted with solid contours to read the effect of including $l > 2$ modes in templates. In both $2-$D and $1-$D panels, solid colored lines mark the true injected parameter value, and dashed vertical lines show the limits of 90% credible regions for the relevant parameter. In all $1-$D panels, the black curve shows the sampling prior for that parameter. See text for further discussion.

regime where PN results are valid, all the way up till the two BHs merge. After merger, the ringdown waveform is constructed as a linear superposition of the first eight quasi-normal modes (QNMs) of the Kerr BH formed at merger [48]. This ringdown waveform is suitably matched with the inspiral-merger portion by enforcing continuity of waveform modes and their first time-derivatives.

We use the most recent SEOBNRv4 model [49] (available within the LIGO Algorithms Library (LAL) [50]) in this study. This model describes BBHs with component spins parallel to the orbital angular momentum (i.e., non-precessing binaries), on non-eccentric orbits, and was calibrated to 141 NR simulations. We refer the reader to [49] and references therein for a comprehensive description of the model. In the interest of minimizing computational cost, we use the reduced-order model for SEOBNRv4 that was also introduced in [49]. We, however, are unable to use the precessing EOB model of [51] in this study due to its high computational cost.

Phenomenological model: IMRPhenomPv2 is a phenomenological model constructed in the frequency domain that describes GWs emitted by non-eccentric spinning-precessing binaries during their inspiral-merger and ringdown phases [31–33]. It relies on the approximation that a generic precessing-binary inspiral waveform can be obtained by rotating the waveform for an equivalent spin-aligned system in its quadrupole-aligned frame to the inertial frame using time-dependent rotors (c.f. PN theory) [31, 32]. In the quadrupole-aligned frame, the leading order $(\ell, m) = (2, \pm 2)$ modes of the waveform are constructed using the non-precessing IMRPhenomD model [33]. The IMRPhenomD model has a closed form in frequency domain, constructed piecewise in three portions: (i) early inspiral: where both mode amplitude and phasing are given by extensions of PN-theory results; (ii)-(iii) late inspiral and ringdown: where phenomenological ansatzes are taken for waveform amplitude and phasing, and calibrated to enforce high-precision agree-

ment with NR simulations from various numerical relativity groups. Note that IMRPhenomPv2(D) belong to the unique class of models that are both closed-form in the frequency domain *and* describe the complete inspiral-merger-ringdown of spin-precessing binaries. These features are ideally suited for GW searches and parameter estimation, which could require the generation of a large number of waveform templates for each event.

C. Parameter Estimation Methodology

Let us denote the collection of measured parameters that describe a GW signal received from a BBH merger event (including the binary's dynamical and kinematic parameters, and other detector-related parameters[2]) as $\vec{\theta}$. The problem statement for PE is to estimate the probability distribution $p(\vec{\theta})$ for the source binary. Using Bayes' theorem, this posterior probability distribution $p(\vec{\theta})$ given data s from GW detectors containing the signal, and a model for GW signals H can be constructed as

$$p(\vec{\theta}|s,H) = \frac{p(s|\vec{\theta},H)\,p(\vec{\theta}|H)}{p(s|H)}, \tag{1}$$

where: (i) $p(s|H)$ is the *prior* expectation of obtaining the new data s, (ii) $p(\vec{\theta}|H)$ is the expectation on parameters $\vec{\theta}$ for astrophysical sources *prior* to obtaining the new data, and (iii) $p(s|\vec{\theta},H)$ is the likelihood of obtaining data s (= signal + noise) given $\vec{\theta}$ describes the signal embedded in it. Assuming the detector noise is stationary colored-Gaussian with zero-mean, we can write

$$p(s|\vec{\theta},H) \propto \exp(-\frac{1}{2}\langle s-h_H(\vec{\theta})|s-h_H(\vec{\theta})\rangle), \tag{2}$$

where $h_H(\vec{\theta})$ is the signal waveform generated with the chosen GW model H, and the noise-weighted inner product $\langle\cdot|\cdot\rangle$ between a and b is defined as

$$\langle a|b\rangle = 4\Re\int_{f_l}^{f_u}\frac{a(f)b^*(f)}{S_n(f)}\mathrm{d}f, \tag{3}$$

with $S_n(f)$ representing the one-sided power spectral density (PSD) of detector noise for LIGO. In this study, we use f_u as the Nyquist frequency corresponding to a sampling rate of 4096Hz. We use the zero-detuning high power design sensitivity curve for Advanced LIGO [54, 55] when not using detector data, and use $f_l = 20$Hz as the lower frequency cutoff. For both events (GW150914 and GW170104) we use LIGO data from its open science center [56], and estimate detector PSD using 1024 seconds of data around the signal concerned as described in [9].

We compute $p(\vec{\theta}|s,H)$ using the Bayesian inference package LALInference [9] that is available as part of the LALSuite software library [50]. LALInference has been extensively used in past analyses published by the LIGO-Virgo Collaborations [1–5], and uses the nested sampling algorithm [57] to estimate source parameters from GW data. We refer the reader to [9] for details of its implementation. As was its original purpose, nested sampling already computes the integrated evidence $Z \equiv p(s|H)$ of the model H. While unimportant to the parameter estimation problem, Z is the key quantity of interest for the purpose of model selection.

For all analyses in this article, we choose sampling priors $p(\vec{\theta}|H)$ identical to those chosen in recent LIGO-Virgo results papers [4, 37], i.e., both BH masses and spin magnitudes are sampled uniformly over their respective ranges, while spin directions are chosen uniformly over a 2−sphere; source distance and sky location are sampled uniformly in 3−D spatial volume out to 2000Mpc, initial inclination angle is sampled uniformly from $[0,\pi]$, and the remaining kinematic parameters are sampled uniformly over their respective ranges. While these priors allow for a direct comparison of our results with published LVC analyses [4, 37], it has been shown [30, 38] that our choice of prior downweights highly spinning binaries for which different choices of prior could improve spin estimation. While neither of these work suggest that GW150914 or GW170104 had large spins, we defer a rigorous study of the effect of priors on the inference of BH spins for these events to future work.

We note from Sec. II A that `NRSur7dq2` is limited to span approximately 40 GW cycles (of the $\ell = |m| = 2$ modes) before merger. Therefore, if the stochastic sampler of LALInference samples a point in binary parameter space for which the complete waveform starting at 20Hz is longer than 40 cycles, the integrated likelihood (c.f. Eq. 3) automatically compensates for the missing cycles by reducing the integrated bandwidth as the surrogate waveform would be zero for frequencies between 20Hz and its start frequency. We do not, however, *a priori* reject such a jump proposal. Finally, we also note that waveform modes included in `NRSur7dq2` templates that have $m > 2$ (such as the $(3,3),(4,4)$ modes) can start at frequencies above 20Hz. In order to mitigate the Gibbs phenomena brought on by the sudden start of these higher-m modes, we taper all templates at their start. However, some of the information in these modes, contained in frequencies between 20Hz and their start, will be ignored in our analyses (as in previous analyses with numerical simulations [23]). We expect though that these modes contribute the most near merger and that the effect of missing lower frequencies should be minimal [23].

[2] such as those that describe instrument calibration uncertainty [52, 53]. For these, we take a conservative estimate of 10%/10° uncertainty in amplitude / phase calibration for both LIGO detectors [53]

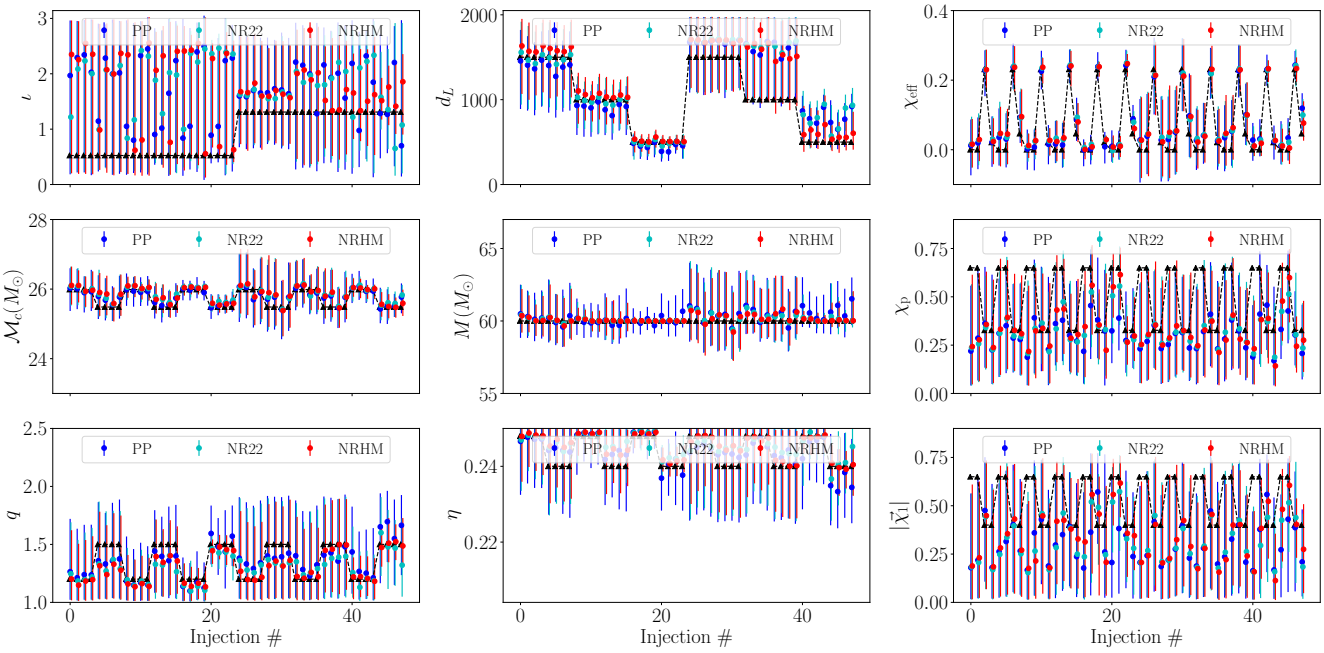

FIG. 2. Estimated masses, spins, and other physical parameters for a set of synthetic GW signals, analyzed using different approximants. In different colors, we show results for 3 approximants: IMRPhenomPv2 (blue, labeled PP), NR22 (`NRSur7dq2` with only $\ell = |m| = 2$ modes) and NRHM (`NRSur7dq2HM`). Details of injected signals are given in the text and Table I. True values of injection parameters are marked by black triangles. Vertical bands show the estimated 90% credible intervals, and colored circles show the estimated median values. The prior range from which mass ratio and spin magnitudes are sampled for IMRPhenomPv2 are identical to the domain of `NRSur7dq2`'s validity.

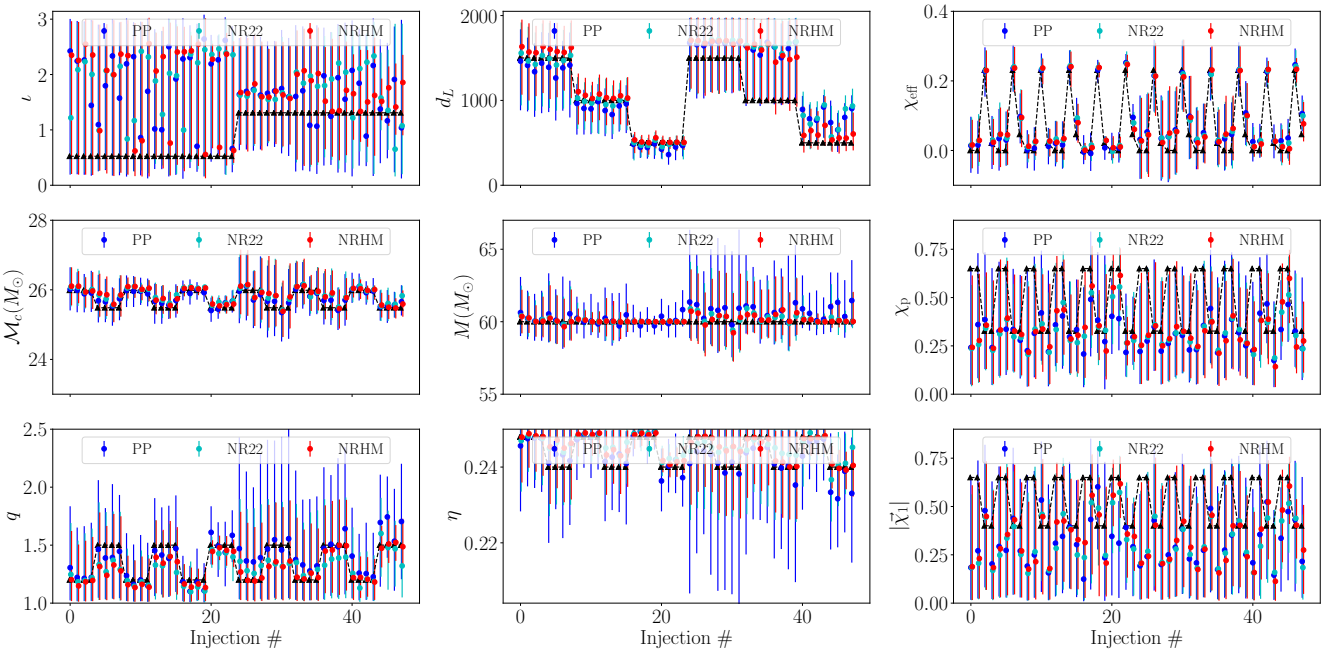

FIG. 3. Same as Fig. 2, except the priors for IMRPhenomPv2 (labeled PP) are widened to span $1 \leq q \leq 8$ and spin magnitudes $|\vec{\chi}_{1,2}| \leq 0.89$.

III. PARAMETER ESTIMATION OF SYNTHETIC GW SIGNALS

As the NRSur7dq2 surrogate model has not been used for GW parameter estimation before, we start with controlled PE tests using synthetic signals to establish its viability for this purpose. Through these tests, we (i) confirm that parameter estimates with NRSur7dq2 are broadly consistent with semi-analytic models; (ii) establish how much and what more information NRSur7dq2 can recover from GW signals; and (iii) measure the loss in information for comparable-mass binaries because of the artificially restricted domain of validity of NRSur7dq2 to mass ratios $1 \leq q \leq 2$ and low to moderately spinning BHs.

We perform a total of 48 injections in zero noise[3] for both LIGO detectors, and analyze the resulting coincident synthetic data with the NRSur7dq2 model (with and without $\ell > 2$ modes), and with the precessing IMRPhenomPv2 model (with and without the artificially restricted priors of NRSur7dq2). We do not include SEOBNRv4 in this section as it only models non-precessing sources, and do not include the precessing EOB model [51] because of its high computational cost of generation [59]. We use the design noise curve for Advanced LIGO detectors [54, 55] while filtering, with a lower frequency cutoff of 20Hz. For the analyses to be relevant for heavy BH binaries, we fix the total mass to be $M = 60M_\odot$ for all signals, and fix all other parameters in the following manner. Mass ratio is chosen from two values $q = \{1.2, 1.5\}$. BH spins are sampled from 4 different configurations: (i) both spins with magnitude 0.65 and both initially parallel to orbital plane with $\hat{\chi}_1 = -\hat{\chi}_2$; (ii) spin on bigger BH with magnitude 0.65 and parallel to orbital plane, and $\vec{\chi}_2 = 0$; (iii) both spins with magnitude 0.4 and mutually parallel with $\hat{\chi}_1 = \hat{\chi}_2 = \frac{1}{\sqrt{3}}(1,1,1)$; and (iv) both spins with magnitude 0.4 and anti-parallel with $\hat{\chi}_1 = -\hat{\chi}_2 = \frac{1}{\sqrt{3}}(1,1,1)$. These spin configurations are chosen to enhance the effects of spin-induced orbital precession. The initial inclination of binary's total angular momentum with the detectors' line of sight is chosen from two values, one close to nearly face-on with $\theta_{\rm JN} = \pi/6$ and another close to edge-on with $\theta_{\rm JN} = 5\pi/12$. Each simulated source is placed at a luminosity distance $d_L = \{500, 1000, 1500\}$Mpc from the detectors. Together these choices form a grid of $2 \times 4 \times 2 \times 3 = 48$ injections. These parameter choices are summarized in Table I. Sources' sky location is chosen from a uniform distribution on a 2−sphere, while the polarization angle, i.e. the third Euler angle required to rotate from the source to detector frame, is chosen from a uniform distribution over $[0, \pi]$. Choices for these angles are held fixed for all injections. We remind the reader that the choice of injected parameters here is made to enhance the effect of precession, and so could be considered a sample of "moderately" to "highly" precessing sources. It is not, however, drawn from an astrophysically motivated distribution, and is therefore *not* representative of a complete astrophysical BBH population. In either case, a sample size of 48 over an 8-D parameter space is unlikely to be *statistically* representative of any chosen distribution. All synthetic signals are generated with the most accurate model available, i.e., NRSur7dq2HM including all $\ell \geq 2$ modes. Note that the choice of using zero noise instead of a particular noise realization ensures that our results hold *on average*, where the averaging is meant in the sense of an ensemble average over an infinite set of noise realizations embedding the same signal. When real instrument noise is present, these results will get shifted depending on the exact nature of the noise realization.

Parameters are sampled and their posterior probability distributions are estimated as described in Sec. II C. For a pedagogical overview, we start with examining a few select injections and study their parameter recovery by NRSur7dq2(HM). We choose 3 injections corresponding to the same binary, with $M = 60M_\odot$, $q = 1.2$, $\vec{\chi}_1 = -\vec{\chi}_2 = \frac{0.65}{\sqrt{2}}(1,1,0)$, located at distances of 1500Mpc, 1000Mpc, and 500Mpc. All of these 3 are inclined to the line of sight at an angle $\theta_{\rm JN} = \pi/6$. We also consider a fourth injection that has the same physical parameters as the rest, is located at 1000Mpc, but is inclined at $\theta_{\rm JN} = 5\pi/12$. For all four, we show the recovery of their mass parameters, and source inclination and distance, in Fig. 1, where they are labeled #0, 8, 16, 32 respectively[4]. Solid lines in all panels showing 1-D histograms indicate the true injected value of the relevant parameter, and dashed vertical lines indicate the bounds of 90% credible intervals. Let us focus first on the left panel of this figure. Looking at chirp mass recovery for the first 3 cases, we notice a stark reduction in the width of measured 90% credible intervals with increasing SNR (or decreasing distance). The measurement of mass ratio also improves as the source moves closer from 1500 to 500Mpc, however more slowly than chirp mass. The fourth injection (#32, shown in cyan) is nearly edge-on with respect to the line of sight for LIGO detectors, with parameters otherwise identical to #8. Comparing it with the others, we immediately see how increasing the source's inclination angle toward $\frac{\pi}{2}$ makes the measurement of BH masses significantly worse. This is because $\theta_{\rm JN} \to \frac{\pi}{2}$ decreases the contribution of $\ell = |m| = 2$ modes and increases that of other $\ell > 2$ waveform modes, and therefore reduces the overall SNR. In the right panel, we show the recovery of source's inclination and distance from LIGO detectors. As on the left, solid lines in the 2-D panel mark the

[3] "zero noise" implies that data is composed of an injected signal plus zeros. Since detector noise is assumed colored-Gaussian with zero mean, using zero noise with a detector-noise weighted likelihood in Eq. 2 makes our analysis equivalent to the average over an *ensemble* of analyses which use actual noise-realizations [58].

[4] These labels will be better defined later in the context of Fig. 2.

true injected values. The solid 2-D contours are the 90% credible regions that we recover with `NRSur7dq2HM`, while their corresponding dashed contours are what we get with `NRSur7dq2` without $l > 2$ modes. Note that dashed contours are only shown in the right panel, where the effect of $l > 2$ multipoles is most striking. We immediately see that the first three injections, which are nearly face-on, have similar 90% credible regions - each two-lobed around face-on and face-off orientations. This is as we expect since both orientations are degenerate and maximize the contribution of dominant $\ell = |m| = 2$ waveform modes. The presence of $l > 2$ modes in recovery templates restricts the distance-inclination posterior further. Regions of the posterior that underestimate luminosity distance are ruled out more aggressively near $\theta_{\rm JN} \to \theta_{\rm JN}^{\rm true}$ than $\theta_{\rm JN} \to \pi - \theta_{\rm JN}^{\rm true}$ for this fiducial binary. We find that this asymmetry between face-on/face-off posterior lobes for distance and inclination also depends on the intrinsic parameters of the source, which modulate the relative signal power content between $l = 3$ and $l = 2$ modes. Further, for $\theta_{\rm JN}$ close to $\pi/2$, we expect a systematic overestimation of distance as the true value can only be located at the lower "U" end of the $d_L - \theta_{\rm JN}$ degeneracy region. We find, accordingly, that distance for the fourth injection is grossly overestimated, in contrast with the other three cases. This is consistent with past results for highly inclined binaries [58].

So far we have illustrated with select cases how measurement errors are reflected in measured posteriors, and consequently in their statistical properties. Next, we will have to analyze a larger number of injections together. We will quantify our results using marginalized 1-D 90% credible intervals as measures of statistical error, with the estimated median values furnishing any corresponding systematic errors. Our results are shown in Fig. 2 for all injections together. In all panels, the horizontal axis shows the injection index, which ranges from $0-47$. Injections are arranged first according to their inclination angle, then according to their source distance, then mass ratio, and finally by bigger BH's spin magnitude. This implies that the first 24 injections shown have source $\theta_{\rm JN} = \pi/6$ and the next 24 have $\theta_{\rm JN} = 5\pi/12$. Within each of these blocks of 24 injections, the first 8 have sources at $d_L = 1500$Mpc, next 8 at $d_L = 1000$Mpc and the last 8 at $d_L = 500$Mpc. Within each of these blocks of 8 injections, the first 4 have $q = 1.2$ and the next 4 have $q = 1.5$. And finally, within each block of 4 injections, the first 2 have $|\vec{\chi}|_1 = 0.65$ and the next 2 have $|\vec{\chi}|_1 = 0.4$. This arrangement shows up in the locations of black triangles in each of the panels in Fig. 2.

In all panels of Fig. 2, IMRPhenomPv2 templates are restricted to sample from the same prior range for mass ratio and component spins that restrict `NRSur7dq2`. Each panel also includes results for `NRSur7dq2HM` with all available $l \leq 4$ modes, and `NRSur7dq2` which includes only the dominant $\ell = |m| = 2$ modes. In the top row of this figure, the left two panels show the recovery of source inclination and distance with respect to the detectors. We know that the effect of these parameters on GW signals received on Earth is degenerate, i.e., signals with inclination $\theta_{\rm JN}$ are degenerate with sources with $\theta_{\rm JN} \to \pi - \theta_{\rm JN}$ at similar distances, as well as with sources having $\theta_{\rm JN} \to \pi/2$ at smaller distances. The general shape of this degeneracy is visually appreciable from the cyan contour in the lower left corner of the right panel in Fig. 1. Therefore, we find that median $\theta_{\rm JN}$ estimates are either close to the true value of $\theta_{\rm JN}$ or π minus the true value. In most cases, neither IMRPhenomPv2 nor `NRSur7dq2` constrain the inclination very well, with the 90% credible intervals spanning the entire prior range $[0, \pi]$. For sources that are close to being edge-on (i.e. with $\theta_{\rm JN} = 5\pi/12$), the effect of higher-order waveform modes is enhanced, and therefore we find that `NRSur7dq2HM` constrains $\theta_{\rm JN}$ better than models with only $\ell = |m| = 2$ modes. This is especially noticeable for sources closer than 1Gpc, (c.f. injections $\#32 - 47$). In the top middle panel we show distance measurements, and from them we notice immediately that distance estimates improve significantly for closer sources, as is expected. For sources with $\theta_{\rm JN} = \pi/6$, we find that `NRSur7dq2HM` measures source distance more accurately and precisely, with both the median estimate being closer to the true value and 90% credible intervals being smaller. For edge-on signals, we find that d_L is *systematically biased* toward larger values, as is expected given the nature of $d_L - \theta_{\rm JN}$ degeneracy. Even then, `NRSur7dq2HM` estimates are both more accurate and more precise for all injections, especially note the loudest ones: $\#40 - 47$.

Next we look at the recovery of the binary's mass parameters. We show these for four different mass combinations - chirp mass $\mathcal{M}_c = M\eta^{3/5}$, total mass M, mass ratio q, and dimensionless mass ratio $\eta = m_1 m_2/M^2$, in the left two columns of the bottom two rows. Starting with chirp mass, we immediately notice that in almost all cases it is constrained to the same degree by all models, categorically worse for nearly edge-on cases than nearly face-on ones. Similar is the case for the estimation of total mass, except that when the sources are closer, both `NRSur7dq2` and `NRSur7dq2HM` estimate M better than IMRPhenomPv2. This pattern holds for nearly edge-on cases as well as for near face-on ones. A similar pattern is seen for the estimation of mass ratio q (or η[5]). While mass ratios are recovered better for nearly face-on systems, they are better estimated by `NRSur7dq2HM` templates than either `NRSur7dq2` or IMRPhenomPv2 ones. This is especially evident for nearly edge-on signals at 500Mpc, i.e. injections $\#40 - 47$.

Next, we consider BH spins. We show the recovery of three different combinations of both BH's spins in the right column (all rows) of Fig. 2: effective spin $\chi_{\rm eff}$, in-plane precessing spin χ_p, and spin magnitude of the big-

[5] the one-to-one map: $\eta = q/(1+q)^2$ ensures that patterns that hold for posteriors of q will hold for η, and vice-versa

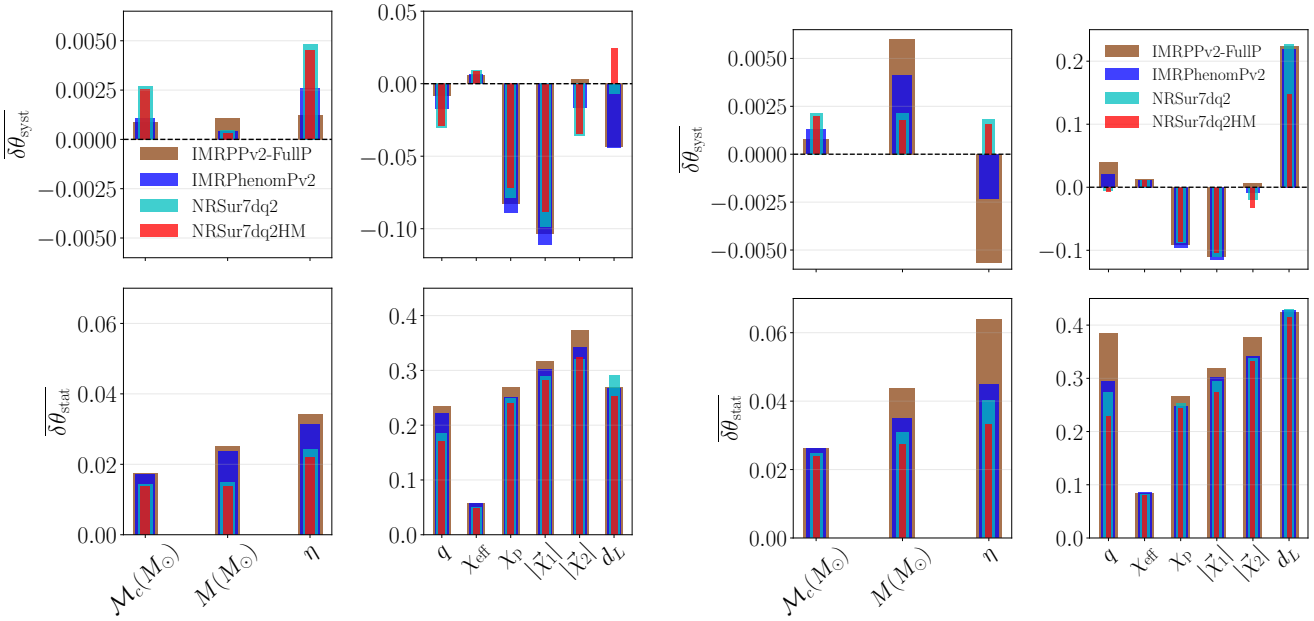

FIG. 4. *Left*: Shown are the mean systematic biases $\overline{\delta\theta_{\rm syst}}$ and statistical uncertainties $\overline{\delta\theta_{\rm stat}}$ for various binary parameters averaged over all injections with $\theta_{\rm JN} = \pi/6$. Four distinct template configurations are considered: the `NRSur7dq2` model with and without $\ell > 2$ modes, and IMRPhenomPv2 with and without being artificially restricted to the validity domain of `NRSur7dq2`. The full prior for IMRPhenomPv2 extends over $1 \leq q \leq 8$ and spin magnitudes $|\vec{\chi}_{1,2}| \leq 0.89$. Since the signal model is `NRSur7dq2HM`, when the recovery model is also `NRSur7dq2HM` the mean systematic biases and statistical uncertainties reflect the shape of the posterior itself rather than modeling error. *Right*: Same as left panel, except the averaging of $\overline{\delta\theta_{\rm syst}}$ and $\overline{\delta\theta_{\rm stat}}$ is performed over all injections with $\theta_{\rm JN} = 5\pi/12$.

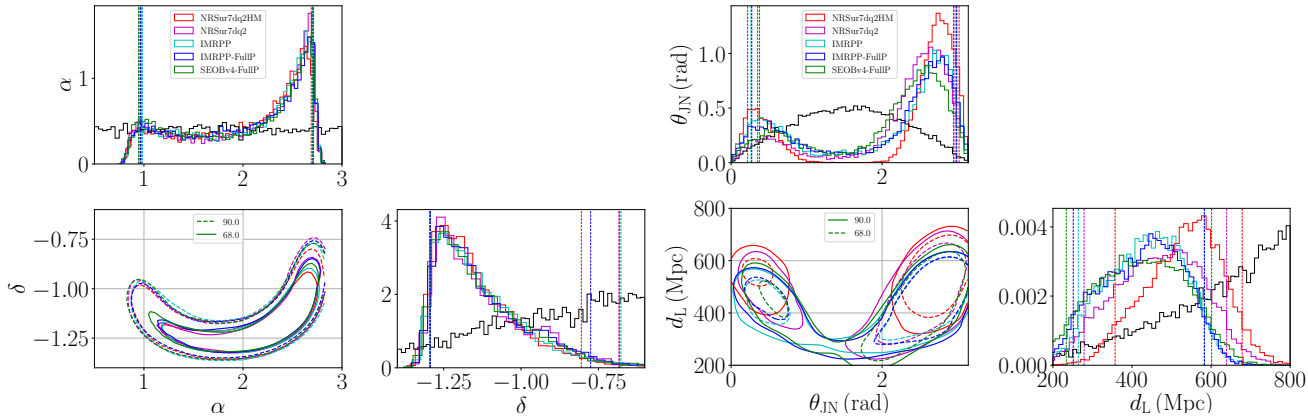

FIG. 5. Estimated sky location (left panel) and orientation / distance (right panel) for GW150914, using different approximants: NRSur7dq2HM, NRSur7dq2, IMRPhenomPv2 (labeled IMRPP), and SEOBNRv4 (labeled SEOBv4). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for IMRPhenomPv2 and $a_{1,2} \lesssim 0.98$ for SEOBNRv4. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1-D posterior distributions, the black line shows our prior belief. Vertical dashed lines in 1-D posteriors mark 90% credible regions. The 2-D posteriors show both the 90% (dashed line) and 68% (solid line) credible regions.

ger BH $|\vec{\chi}_1|$ (from top to bottom). From the top panel we note that effective spin is estimated well by all template models. For the closest sources (at 500Mpc) it is somewhat better estimated by `NRSur7dq2HM`. From the middle right and bottom right panels, we note that both χ_p and $|\vec{\chi}_1|$ are poorly constrained for heavy BBHs such as the cases considered, and their 90% credible regions almost span 90% of the entire prior range, implying that very little information about these spin parameters is extractable. This is to be expected because the timescale

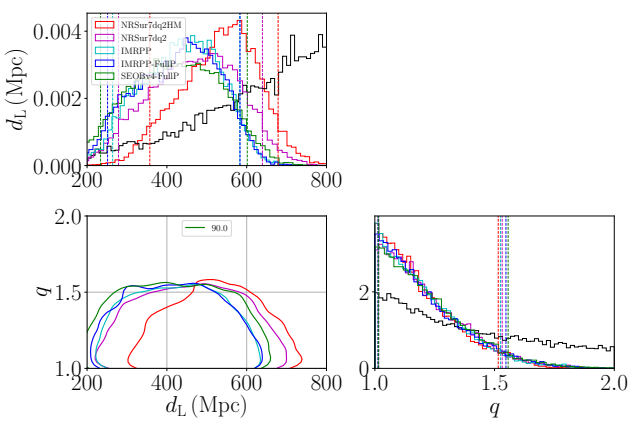

FIG. 6. luminosity distance and mass ratio measurement for GW150914. All figure attributes are similar to Fig. 5. We find that the samples at large luminosity distances actually correspond to *smaller* mass ratios, and therefore the shifting of distance measurement to larger values when using `NRSur7dq2HM` is not a symptom of the model's restricted sampling priors.

of orbital precession is much longer than the orbital timescale, and short signals from heavy BBHs barely span a couple of precession cycles, making the measurements of precession-related parameters challenging. We still note, however, that for the closest injected sources, `NRSur7dq2HM` recovers $|\vec{\chi}_1|$ more accurately than either `NRSur7dq2` or IMRPhenomPv2, indicating that harnessing the signal power in sub-dominant GW modes could help some of our spin measurements.

Next, in Fig. 3 we show similar results with the only difference from Fig. 2 being that the IMRPhenomPv2 priors are no longer restricted to the `NRSur7dq2` domain. Instead, samples are drawn from $q \in [1, 8]$ and spin magnitudes from $[0, 0.89]$. Comparing these two figures demonstrates the effect of `NRSur7dq2`'s limited priors on the parameter recovery results discussed so far. We notice that mass ratios are naturally less constrained when IMRPhenomPv2 is allowed to wander over a larger range, especially for edge-on systems, but other than that, the estimation of other mass / spin / distance / inclination parameters is remarkably similar to the previous results. This demonstrates that the effect of priors on our results is minimal.

One way to summarize the information from all injections is to compute an averaged measure of systematic biases and statistical uncertainties associated with the recovery of various physical parameters θ by different waveform models. We compute these averages by first computing the relative systematic bias $\delta\theta^i_{\rm syst}$ and relative statistical uncertainty $\delta\theta^i_{\rm syst}$ for injection i:

$$\begin{aligned} \delta\theta^i_{\rm syst} &:= \left|\theta^i_{\rm median} - \theta^i_{\rm true}\right| / \theta^i_{\rm true} \\ \delta\theta^i_{\rm stat} &:= \left|\Delta\theta^i_{90\%}\right| / \theta^i_{\rm true}, \end{aligned} \tag{4}$$

where $\Delta\theta^i_{90\%}$ is the size of the 90% credible region. For parameters whose possible values include 0, such as BH spins and their combinations, we do not divide by $\theta^i_{\rm true}$ in both parts of Eq. 4. Subsequently we take the algebraic mean of both $\delta\theta^i_{\rm syst}$ and $\delta\theta^i_{\rm stat}$ over all 48 injections to obtain our combined measures of parameter estimation accuracy and precision: $\overline{\delta\theta_{\rm syst}}$ and $\overline{\delta\theta_{\rm stat}}$. We find that both of these measures are significantly affected by the inclination angles of injections being averaged over. We therefore average over two sets of injections separately, one where the injected $\theta_{\rm JN} = \pi/6$ and another where $\theta_{\rm JN} = 5\pi/12$.

Both summary error measures $\overline{\delta\theta_{\rm syst}}$ and $\overline{\delta\theta_{\rm stat}}$ are shown in Fig. 4. Lets first focus on the left panel, bottom sub-panels. We find that the surrogates recover all mass combinations - $\mathcal{M}_c, M, \eta, q$ - *with more precision* than IMRPhenomPv2. The difference that restricting the prior makes to the IMRPhenomPv2 analysis is smaller in comparison (seen by comparing the dark blue and brown bars in the graphic), and therefore the improved precision with `NRSur7dq2HM` is unlikely to be due to its restricted domain of validity. That `NRSur7dq2` also performs similarly suggests that this improvement is due to the improved modeling of the dominant waveform mode. Next, we look at the systematic biases in the recovery of the same mass parameters shown in the sub-panels right above. We find them to be between $1-10\%$ of the corresponding statistical errors. They are therefore negligible as the latter dominates our mass measurement. Further, we look at various spin combinations - $\chi_{\rm eff}, \chi_{\rm p}, |\vec{\chi}_1|, |\vec{\chi}_2|$ - to the right of mass parameters in the left panel of Fig. 4. As before, we look at the statistical uncertainties first. We find that both `NRSur7dq2(HM)` surrogates allow for slight reduction in statistical errors for all spin combinations, including $\chi_{\rm eff}$. In the sub-panel right above, we can read the systematic biases in measuring these combinations as well. We find that while for $\chi_{\rm p}$ and $|\vec{\chi}_1|$ the surrogates record smaller systematic biases in measurement, for $\chi_{\rm eff}$ and the smaller BH's spin its the opposite. However, for both $\chi_{\rm eff}$ and $|\vec{\chi}_2|$ the ratio $\overline{\delta\theta_{\rm syst}}/\overline{\delta\theta_{\rm stat}}$ is below 10% and the measurement is dominated by statistical errors. Having said this, we remind the reader that for most spin combinations, except for $\chi_{\rm eff}$, little information is actually recovered from data and measurements tend to follow the prior distributions. Next, we assess the errors in luminosity distance measurement. Looking at the lower right sub-panel first, we find that both surrogates and IMRPhenomPv2 measure d_L with comparable precision. From the sub-panel right above, we find that the `NRSur7dq2HM` systematically over-estimates d_L by about 2.5% while IMRPhenomPv2 under-estimates it by 5%. We know that both $d_L - \theta_{\rm JN}$ are degenerate in how they effect the signal received by GW detectors. This degeneracy shows up as correlations in posterior distributions, for e.g. see the cyan contour in the bottom left sub-panel of the right panel of Fig. 1. Sources that are nearly face-on or face-off are correlated with those that are nearly edge-on but are much closer. The inclusion of higher-order modes in `NRSur7dq2HM` can make this degeneracy resolvable. For nearly face-on sources

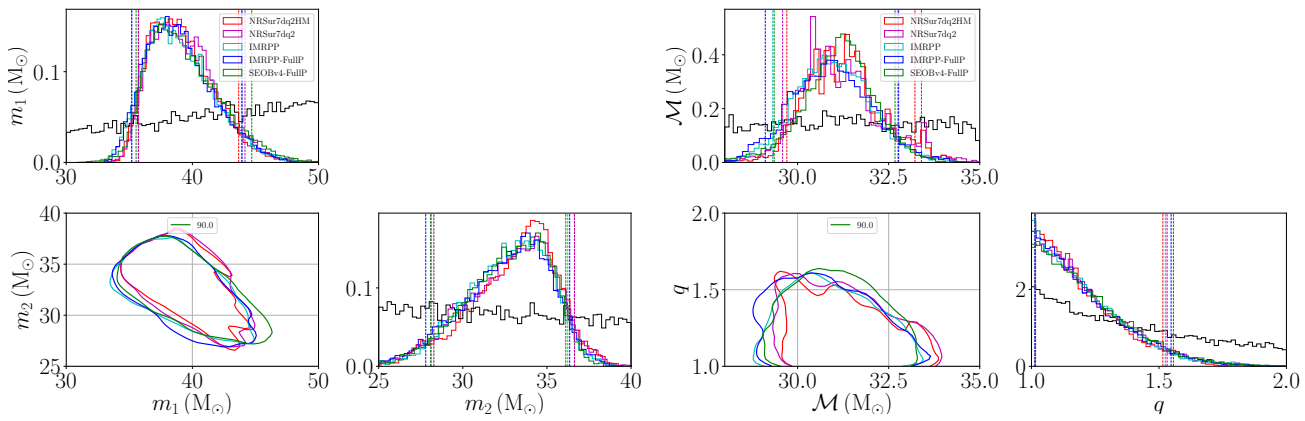

FIG. 7. Estimated masses for GW150914, using different approximants: NRSur7dq2HM, NRSur7dq2, IMRPhenomPv2 (labeled IMRPP), and SEOBNRv4 (labeled SEOBv4). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for IMRPhenomPv2 and $a_{1,2} \lesssim 0.98$ for SEOBNRv4. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1-D posterior distributions, the black line shows our prior belief. Vertical dashed lines in 1-D posteriors mark 90% credible regions. The 2-D posteriors show the 90% credible regions as a solid line.

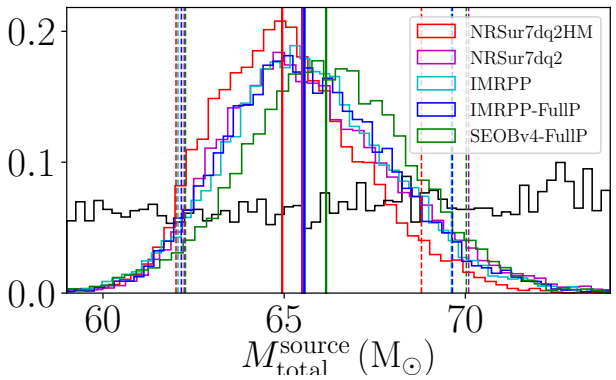

FIG. 8. Estimated total mass for GW150914, as measured in the source frame. Four different approximants are shown: NRSur7dq2HM, NRSur7dq2, IMRPhenomPv2 (labeled IMRPP), and SEOBNRv4 (labeled SEOBv4). Figure attributes are identical to Fig. 7. The black line shows our prior belief. Vertical dashed lines mark 90% credible regions, and vertical solid lines show the distribution median.

such as the injections the entire left panel of Fig. 4 is averaged over, we find that the lower-mass-nearly-edge-on region of the $d_L - \theta_{\mathrm{JN}}$ posterior gets ruled out by `NRSur7dq2HM` and the entire marginalized posterior for d_L moves to larger values. This suggests that distance estimates with `NRSur7dq2HM` would be *systematically larger* than those with either `NRSur7dq2` or IMRPhenomPv2, and is consistent with $\overline{\delta\theta_{\mathrm{syst}}}$ being strictly positive for d_L with `NRSur7dq2HM`. However, the ratio $\overline{\delta\theta_{\mathrm{syst}}}/\overline{\delta\theta_{\mathrm{stat}}}$ for d_L is 10% with `NRSur7dq2HM`, 4% for `NRSur7dq2`, and 16% for IMRPhenomPv2, suggesting that systematic errors in measuring d_L are small in comparison to statistical errors and therefore the latter dominates for these injections.

We conclude that `NRSur7dq2HM` could both improve the accuracy and precision (somewhat) of d_L measurement. Moving forward, we focus on the edge-on injections in the right panel of Fig. 4. Compared to the left panel, we immediately note that the measurement of M, η, q is dominated by the restriction on the domain of validity of `NRSur7dq2HM`. For χ_{eff}, again, both surrogates and IMRPhenomPv2 perform comparably, while for other spin combinations all effectively recover the sampling prior. Finally, we note that for edge-on injections the d_L is systematically *over*-estimated by all models, as has been found before [58] (see Fig. 4). Even then, the surrogate furnishes a 30% more accurate measurement due to the inclusion of sub-dominant modes.

From these results, we conclude that the results we obtain with `NRSur7dq2HM` templates can be an improvement over those coming from the precessing waveform model that has been used so far to analyze LIGO-Virgo BBH observations. We also conclude that for comparable mass heavy close-by sources ($\sim$ 500Mpc) and/or nearly edge-on sources, using the `NRSur7dq2` surrogate model with all of its available waveform modes could improve the estimation of the binary's luminosity distance (and therefore inclination angle) and effective spin. Our results emphasize the impact of accurate templates with higher harmonics in parameter estimation. This is consistent with past work on higher harmonics [34, 35, 60, 61]. Based on these results, we encourage the community to utilize NR surrogates for detailed follow-ups of heavy BBH coalescences.

IV. GW150914

Having assessed the performance of `NRSur7dq2` surrogate models in characterizing gravitational-wave signal

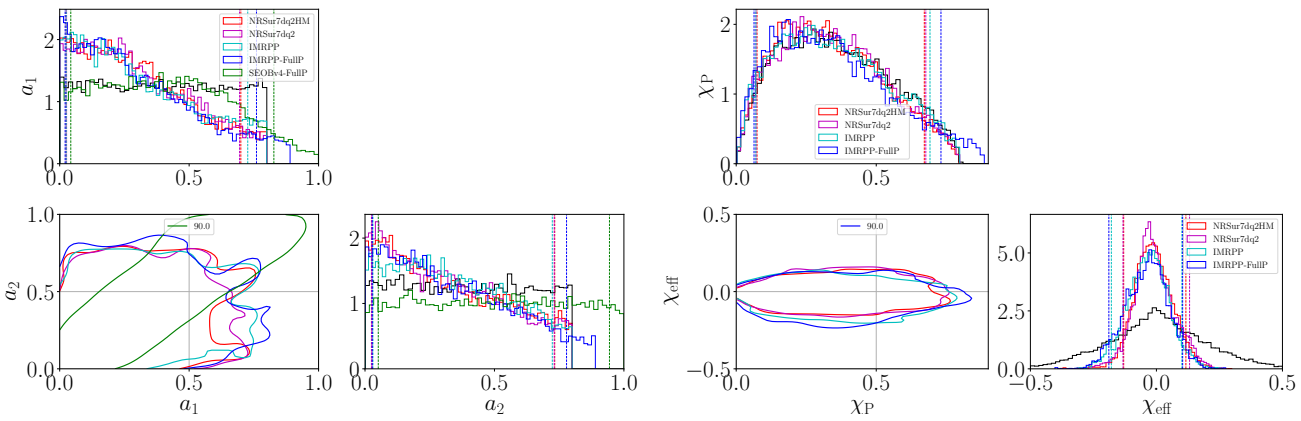

FIG. 9. Estimated spins for GW150914, using different approximants and different prior probability distributions. Shown are spin magnitudes for both BHs and the tilt angles between BH spins and the orbital angular momentum at $f_{\rm ref}$. Figure attributes are identical to Figs. 5 and 7.

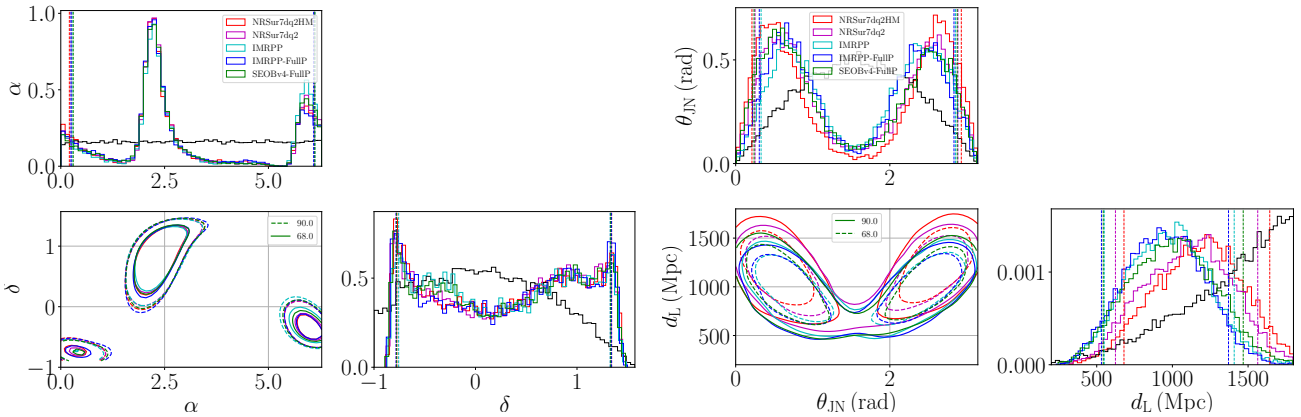

FIG. 10. Estimated sky location (left panel) and orientation / distance (right panel) for GW170104, using different approximants: NRSur7dq2HM, NRSur7dq2, IMRPhenomPv2 (labeled IMRPP), and SEOBNRv4 (labeled SEOBv4). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for IMRPhenomPv2 and $a_{1,2} \lesssim 0.98$ for SEOBNRv4. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1-D posterior distributions, the black line shows our prior belief. Vertical dashed lines in 1-D posteriors mark 90% credible regions. The 2-D posteriors show both the 90% (dashed line) and 68% (solid line) credible regions.

sources in a fully Bayesian framework, we now analyze the first ever observed GW event GW150914 with these models. As before, we use the nested sampling algorithm in LALInference to perform parameter recovery on the event, and use non-precessing SEOBNRv4 templates in addition to `NRSur7dq2` and IMRPhenomPv2. We perform two analyses with IMRPhenomPv2: one where we artificially restrict sampling priors to the domain of `NRSur7dq2` and another where we do not. We, however, find that both analyses furnish almost identical results, and therefore conclude that the effect of sampling priors on GW150914's analyses is minimal. In the analysis with SEOBNRv4, we do not artificially restrict the sampling prior. As in the previous section, we do not use the precessing EOB model of [51] due to its high computational cost. Results from all of the above analyses are posterior probability distributions for physical parameters describing the GW source, which are shown in Figs. 5 - 9. In all of these figures, black curves in 1-D posterior distributions will show prior distributions for respective parameters.

In Fig. 5, left panel, we show the recovery of the source's sky location angles (right ascension α and declination δ) by `NRSur7dq2` and compare it with those for semi-analytic models. We immediately note that the recovery of sky location of GW150914's source with `NRSur7dq2` is remarkably similar to that from semi-analytic models, but adds little extra information. In the right panel of the same figure, we show the recovery of the source's luminosity distance d_L from LIGO detectors and its total angular momentum's initial inclination $\theta_{\rm JN}$ with respect to the line of sight. These two parameters are

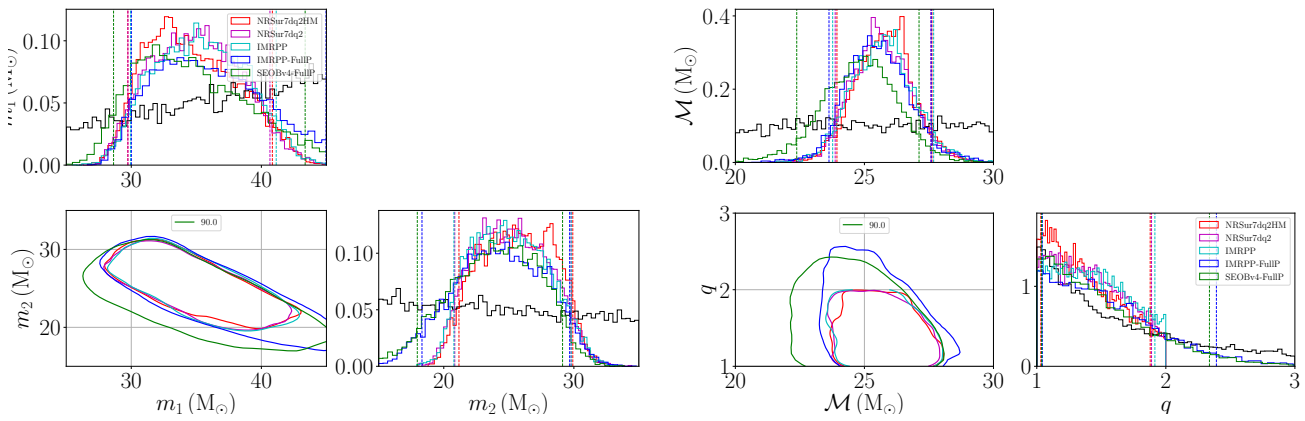

FIG. 11. Estimated masses for GW170104, using different approximants: NRSur7dq2HM, NRSur7dq2 and IMRPhenomPv2 (labeled IMRPP). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for IMRPhenomPv2 and $a_{1,2} \lesssim 0.98$ for SEOBNRv4. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1-D posterior distributions, the black line shows our prior belief. Vertical dashed lines in 1-D posteriors mark 90% credible regions. The 2-D posteriors show the 90% credible regions as a solid line.

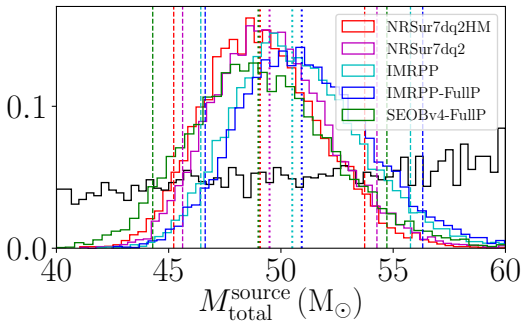

FIG. 12. Total mass for GW170104, as measured in the source frame. Four different approximants are shown: NRSur7dq2HM, NRSur7dq2, IMRPhenomPv2 (labeled IMRPP), and SEOBNRv4 (labeled SEOBv4). All figure attributes are identical to the previous figure.

strongly degenerate, as can be seen from the 2-D posterior slices showing 1-σ and 90% credible regions for both. We also note the effect of higher-order waveform multipoles included in `NRSur7dq2HM` on the measurement of both d_L and θ_{JN}. From their 90% credible intervals in 1-D marginalized posteriors, we can see that `NRSur7dq2HM` places GW150914 at $\sim$ 530Mpc, while other models, including `NRSur7dq2`, place it at $\sim$ 430Mpc. The primary LVC analyses of the event also inferred the source to be at $\sim$ 410Mpc [37]. This strongly implies that the difference in luminosity distance estimation is indeed due to the inclusion of higher-order waveform modes. Similarly, the inclination angle is more precisely constrained by `NRSur7dq2HM` to be either face-on or face-off, with edge-on configurations being more strongly disfavored by it than all other models. *These are some of the key findings of this paper. They were inaccessible to the original LVC analyses [4, 23, 37], which were limited by modeling approximations and the availability of a sufficient number of NR simulations.* In Fig. 6 we show the correlated posterior distribution for luminosity distance and mass ratio. We see immediately that the increase in the estimated value of d_L by `NRSur7dq2HM` is not an artifact of the model's restricted domain of validity since the region of the posterior at large distances actually corresponds to nearly equal-mass binaries.

Next, we show the recovery of mass parameters for GW150914 in Fig. 7. The left panel shows 1-D and 2-D marginalized posteriors for individual BH masses, and the right panel shows the same for the binary's chirp mass and mass ratio. While for individual masses and mass ratio, `NRSur7dq2HM` and `NRSur7dq2` give us very similar posterior distributions to what we obtain from approximate waveform models, the binary's chirp mass is estimated to be somewhat higher by both surrogate models. While this difference is marginal, it is consistent with `NRSur7dq2HM`'s estimation of luminosity distance to larger values, as the GW signal strength depends on the ratio of the two. Further, given that the distance estimated by `NRSur7dq2HM` and other waveform models differs by $\sim 15\%$, we expect the measured source-frame mass to also differ by $\delta M^{\mathrm{source}} \sim -M\delta z$ where δz is the corresponding difference in the inferred redshift of source (assuming standard cosmology [62]) between models, and M is the estimated total mass. This can be seen from Fig. 8 where we show the posterior distribution for the total mass of the binary in its source frame. We find that `NRSur7dq2HM`'s median estimate to be approximately $0.5 - 1M_\odot$ *lower* than others, which is consistent with our estimate of $-0.15 \times 0.1 \times 65 \approx -1M_\odot$ (see also Table II).

Finally, we focus on the recovery of binary spins for GW150914 in Fig. 9. The left panel shows marginal-

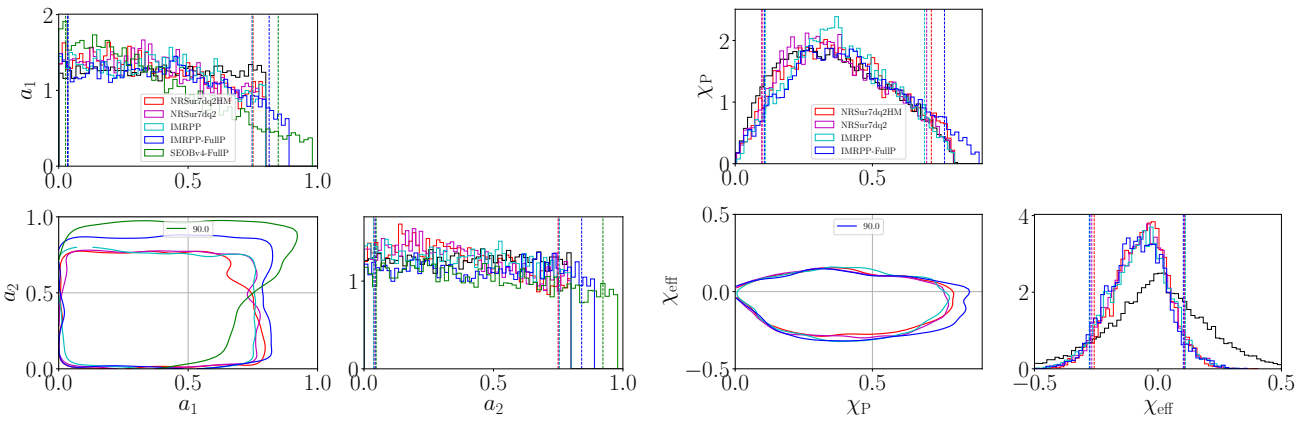

FIG. 13. Estimated spins for GW170104, using different approximants and different prior probability distributions. Shown are spin magnitudes for both BHs and the tilt angles between BH spins and the orbital angular momentum at $f_{\rm ref}$. Figure attributes are identical to Figs. 10 and 11.

ized 1-D and 2-D posteriors for individual BH spin magnitudes (labeled $a_{1,2} \equiv |\vec{\chi}_{1,2}|$), and the right panel focuses on their effective-spin $\chi_{\rm eff} := (m_1\chi_{1z} + m_2\chi_{2z})/M$ and precessing-spin χ_p [32] combinations. From the left panel we note that spin recovery with `NRSur7dq2HM` and `NRSur7dq2` closely follows what we measure with our approximate precessing model IMRPhenomPv2. From SEOBNRv4 we find spin magnitudes to be constrained along the region with $|\vec{\chi}_1| \simeq |\vec{\chi}_2|$. This is as expected given that the effective spin combination is constrained close to 0 (right panel), which necessitates $\vec{\chi}_1 \simeq -\vec{\chi}_2$ for a comparable-mass binary such as this. From the right panel, we note that the surrogate model does not recover any additional information about the binary's precessing spin component, as its posterior appears to be sampling the prior with little information being added by data. However, *it does constrain the source's effective-spin to be somewhat closer to zero.* As both `NRSur7dq2HM` and `NRSur7dq2` provide for better estimation of $\chi_{\rm eff}$, we conclude that this may be because of additional spin information in the surrogate models that is not included in the IMRPhenomPv2 model.

Overall, we conclude that *the use of* `NRSur7dq2` *surrogate improves the estimation of source distance and inclination for GW150914 because of the inclusion of higher-order modes, and helps constrain the binary's effective spin better because of capturing the nonlinearities of GR better than approximate waveform models.* These results are further quantified in Table II, which can be directly compared with Table I of [37].

V. GW170104

The second heavy binary black hole merger was detected by the two LIGO detectors on January 4, 2017[6]. We perform identical analyses on this event as we did for GW150914. Their results are shown in Figs. 10 - 13.

In the left panel of Fig. 10 we show the recovery of GW170104's sky location with different models and choices of priors. We immediately note that all models yield remarkably similar estimates for the sky location, with the inclusion of higher-order modes in `NRSur7dq2HM` not yielding much additional information. The right panel of the same figure shows the recovery of luminosity distance d_L and initial inclination angle $\theta_{\rm JN}$ with respect to the line of sight. Qualitatively similar to GW150914, we find that `NRSur7dq2HM` narrows the range of plausible $\theta_{\rm JN}$ values to be closer to face-on and face-off configurations (than edge-on ones). It also estimates the source to be located at a median distance of ~ 1080Mpc which is 20% larger than what we get when using approximate models here (882Mpc) as well as in the published LVC analysis of the event (880Mpc) [4]. $8-10\%$ further away than what is estimated by both precessing and non-precessing models. As was shown for GW150914 in Fig. 6, for GW1701014 too we find that the increase in the estimated value of d_L with `NRSur7dq2HM` is not due to the model's restricted domain of validity.

Next, we show the estimation of binary mass parameters for GW1701014 using different models in Fig. 11. We note that the 2-D posterior distribution for individual BH masses has support at mass ratios larger than $q = 2$ and therefore `NRSur7dq2` models only recover a

[6] The second actual detection was GW151226 [2]. In the context of this paper, this event was a "light" BBH merger. Most of its SNR is contributed by the long inspiral, which is well modeled by semi-analytic approximate waveforms.

fraction of the whole posterior. However, the chirp mass estimated by all precessing models is consistent, while the non-precessing SEOBNRv4 model constrains it less stringently. Overall we find results from semi-analytic models to be consistent with NR surrogate estimates. Similar to GW150914, we expect the source-frame mass of GW170104 to be measured differently by `NRSur7dq2HM` than other waveform models. We confirm this through Fig. 12 where we show the posterior distribution for the total mass of the binary in its source frame. We find that `NRSur7dq2HM`'s median estimate to be approximately $1M_\odot$ lower than others, see Table II for other mass parameters measured in source frame.

In Fig. 13 we demonstrate how well we estimate component BH spins for GW170104 with the NR surrogate model, and compare it with what we get from semi-analytic ones. In the left panel we show the estimation of spin magnitudes, while in the right panel we show the same for the effective spin $\chi_{\rm eff}$ and precessing spin χ_p combinations. While all models estimate $\chi_{\rm eff} \simeq -0.1$ for this event, they recover little information for either $a_{1,2}$ or χ_p, with their respective 1-D posteriors following closely their sampling priors. For all spin combinations considered, we note that the recovery from all models is remarkably similar, despite the additionally restricted priors of `NRSur7dq2` models. For this signal, however, the `NRSur7dq2(HM)` and approximate models provide essentially identical spin information.

Overall, we conclude that with the NR surrogate model `NRSur7dq2` we estimate the source of GW170104 to be approximately 20% further away than was previously estimated using semi-analytic waveform models [4]. The same surrogate furnishes little extra information for source mass and spin parameters of GW170104 though. Our results are summarized in Table II.

VI. DISCUSSION

The population of binary black hole mergers that the LIGO-Virgo detector network has seen so far consists primarily of loud signals coming from "heavy" black hole binaries, with each hole measuring around $20-30$ times the mass of the Sun. Coalescing binaries of such heavy black holes radiate gravitational waves at lower frequencies than their lighter counterparts, and therefore enter the sensitive frequency band of current GW detectors only a short while before they merge. During this pre-merger period of the two-body evolution that is visible to LIGO-Virgo, inspiraling binaries' orbits evolve rapidly from being well approximated as a sequence of slowly shrinking spheres (or circles) to being highly dynamical as both holes enter each other's strong-field regions at highly relativistic velocities. Describing the motion in this pre-merger regime, and consequently the form of emitted gravitational radiation, is beyond the reach of traditional perturbative methods that typically rely on the dynamical timescale of gravity being large and/or binary motion being non-relativistic.

Fully numerical solutions of nonlinear Einstein equations is the most powerful (and only) approach that can tackle the physics in the pre-merger regime. This, however, comes at a non-trivial computational cost that precludes numerical simulations of an arbitrary number of binary configurations. With present day technology and budgets, it is possible to perform approximately $\mathcal{O}(10^3)$ simulations in a calendar year. However, when trying to determine the physical parameters of the source of a BBH merger event from its observed GW data, one typically needs to matched-filter the data against $\mathcal{O}(10^{6-8})$ distinct GW templates[7]. Therefore there is a large gap between the demand of matched-filtering templates and their availability through direct numerical simulations. This gap is traditionally bridged by introducing *phenomenological* extensions to perturbative waveform models, and *calibrating* these extensions to agree with numerical simulations where they can. Examples of such models would include those within the Effective-one-body family [46] and the phenomenological family [18]. Although these models now span a fair region of the full 7-dimensional parameter space that describes arbitrary binary black hole coalescences, there is always scope for inaccuracies in one corner or another [21, 22]. An ab-initio more accurate and reliable approach would be to develop a 7-dimensional numerical interpolant (a *surrogate* model) for the gravitational-wave strain based on select numerical relativity simulations. Such a surrogate model would have been too expensive to construct in the past. With improvements to numerical relativity technology in recent years, Blackman et al [28, 29] developed the first usable surrogate models based solely on numerical relativity simulations. They span a restricted region of the full 7-D binary black hole parameter space, but within that region the model describes arbitrary precessing binary orbits.

The primary purpose of this paper is to use the numerical relativity-based surrogate model `NRSur7dq2` [29] in the fully Bayesian framework and demonstrate both its viability and efficacy for estimating source parameters from heavy binary black hole merger signals. This work paves the way for future surrogate models that will gradually span the entire parameter space for most GW events. We also use the same surrogate [29] to re-analyze data from the first two heavy BBH merger events: GW150914 and GW170104. While we find improvements in the precision of measuring mass and spin parameters for these events' source binaries, our primary finding is that both of these events were located $15-20\%$ further away than what waveform-model based analyses found, including published LIGO-Virgo results [4, 37].

[7] Recently proposed grid-based methods [24, 30] can recover a *subset* of binary parameters with much fewer ($\mathcal{O}(10^3)$) template evaluations. However, the impact of approximations used in these methods on unstructured NR grids (physical and technical) will need to be quantified more thoroughly.

TABLE II. Summary of the parameters that characterize GW150914 and GW170104. For model parameters we report the median value as well as the range of the symmetric 90% credible interval [63]; where useful, we also quote 90% credible bounds. The source redshift and source-frame masses assume standard cosmology [62]. The spin-aligned SEOBNRv4 (labeled SEOB) and precessing IMRPhenomPv2 (labeled IMRPP) waveform models are described in the text, as is the NR surrogate labeled here `NRSur7dq2HM`. Results for the effective precession spin parameter $\chi_{\rm P}$ used in the IMRPP model are not shown as we effectively recover the prior; see left panels of Figs. 9 and 13. The SEOB/IMRPP values stated here are directly comparable to Table I of [37] for GW150914 and to Table I of [4] for GW170104, and are broadly consistent with published LIGO analyses.

	GW150914 SEOB / IMRPP / `NRSur7dq2HM`	GW170104 SEOB / IMRPP / `NRSur7dq2HM`
Detector-frame total mass $M/M_\odot$	$72.1^{+3.5}_{-4.0}$ / $71.6^{+4.1}_{-3.8}$ / $72.2^{+4.8}_{-3.1}$	$58.8^{+5.8}_{-5.4}$ / $60.2^{+5.7}_{-5.3}$ / $59.5^{+4.8}_{-4.6}$
Detector-frame chirp mass $\mathcal{M}/M_\odot$	$31.2^{+1.5}_{-1.8}$ / $30.9^{+1.9}_{-1.8}$ / $31.2^{+2.0}_{-1.5}$	$24.9^{+2.2}_{-2.5}$ / $25.1^{+2.3}_{-3.0}$ / $25.4^{+2.2}_{-2.1}$
Detector-frame primary mass $m_1/M_\odot$	$39.0^{+5.7}_{-3.5}$ / $38.7^{+5.2}_{-3.5}$ / $38.9^{+4.8}_{-3.2}$	$34.3^{+9.1}_{-5.7}$ / $37.0^{+9.7}_{-6.6}$ / $34.1^{+5.9}_{-4.8}$
Detector-frame secondary mass $m_2/M_\odot$	$32.9^{+3.2}_{-4.8}$ / $32.9^{+3.4}_{-5.1}$ / $33.5^{+3.2}_{-5.2}$	$24.1^{+5.1}_{-6.1}$ / $22.7^{+6.3}_{-6.6}$ / $25.4^{+4.5}_{-5.1}$
Detector-frame final mass $M_{\rm f}/M_\odot$	$68.7^{+3.1}_{-3.6}$ / $68.2^{+3.7}_{-3.4}$ / $68.8^{+4.2}_{-2.7}$	$56.3^{+5.4}_{-4.9}$ / $57.8^{+5.7}_{-4.9}$ / $56.9^{+4.4}_{-4.2}$
Source-frame total mass $M^{\rm source}/M_\odot$	$66.2^{+3.2}_{-3.7}$ / $65.5^{+3.8}_{-3.5}$ / $64.9^{+4.3}_{-2.8}$	$49.0^{+4.8}_{-4.5}$ / $51.1^{+4.9}_{-4.5}$ / $49.1^{+3.9}_{-3.8}$
Source-frame chirp mass $\mathcal{M}^{\rm source}/M_\odot$	$28.6^{+1.4}_{-1.7}$ / $28.3^{+1.7}_{-1.7}$ / $28.1^{+1.8}_{-1.4}$	$20.7^{+1.9}_{-2.1}$ / $21.3^{+1.9}_{-2.6}$ / $21.0^{+1.8}_{-1.7}$
Source-frame primary mass $m_1^{\rm source}/M_\odot$	$35.8^{+5.2}_{-3.2}$ / $35.5^{+4.8}_{-3.2}$ / $35.0^{+4.3}_{-2.9}$	$28.6^{+7.6}_{-4.7}$ / $31.4^{+8.2}_{-5.6}$ / $28.1^{+4.8}_{-4.0}$
Source-frame secondary mass $m_2^{\rm source}/M_\odot$	$30.2^{+2.9}_{-4.4}$ / $30.1^{+3.2}_{-4.7}$ / $30.1^{+2.9}_{-4.7}$	$20.1^{+4.2}_{-5.1}$ / $19.3^{+5.3}_{-5.6}$ / $20.9^{+3.7}_{-4.2}$
Source-fame final mass $M_{\rm f}^{\rm source}/M_\odot$	$63.0^{+2.9}_{-3.3}$ / $62.5^{+3.3}_{-3.1}$ / $61.8^{+3.8}_{-2.5}$	$47.0^{+4.5}_{-4.1}$ / $49.1^{+4.8}_{-4.2}$ / $47.0^{+3.6}_{-3.4}$
Mass ratio q	$1.2^{+0.4}_{-0.2}$ / $1.2^{+0.4}_{-0.2}$ / $1.2^{+0.4}_{-0.1}$	$1.4^{+0.9}_{-0.4}$ / $1.6^{+1.1}_{-0.6}$ / $1.3^{+0.6}_{-0.3}$
Effective inspiral spin parameter $\chi_{\rm eff}$	$-0.01^{+0.11}_{-0.15}$ / $-0.03^{+0.14}_{-0.16}$ / $-0.01^{+0.13}_{-0.12}$	$-0.12^{+0.21}_{-0.28}$ / $-0.10^{+0.19}_{-0.23}$ / $-0.09^{+0.19}_{-0.20}$
Dimensionless primary spin magnitude a_1	$0.40^{+0.43}_{-0.36}$ / $0.26^{+0.50}_{-0.24}$ / $0.27^{+0.43}_{-0.25}$	$0.32^{+0.53}_{-0.29}$ / $0.42^{+0.40}_{-0.37}$ / $0.41^{+0.35}_{-0.37}$
Dimensionless secondary spin magnitude a_2	$0.50^{+0.44}_{-0.45}$ / $0.30^{+0.48}_{-0.28}$ / $0.30^{+0.44}_{-0.27}$	$0.45^{+0.47}_{-0.41}$ / $0.46^{+0.39}_{-0.41}$ / $0.39^{+0.37}_{-0.35}$
Final spin $a_{\rm f}$	$0.69^{+0.04}_{-0.07}$ / $0.67^{+0.05}_{-0.06}$ / $0.68^{+0.04}_{-0.05}$	$0.62^{+0.10}_{-0.15}$ / $0.61^{+0.08}_{-0.18}$ / $0.64^{+0.07}_{-0.09}$
Luminosity distance $D_{\rm L}$/Mpc	$423.4^{+178.0}_{-189.7}$ / $434.1^{+149.2}_{-182.7}$ / $538.5^{+140.2}_{-181.3}$	$1000.9^{+467.0}_{-457.9}$ / $882.3^{+407.6}_{-376.9}$ / $1079.3^{+441.3}_{-487.1}$
Source redshift z	$0.09^{+0.03}_{-0.04}$ / $0.09^{+0.03}_{-0.04}$ / $0.11^{+0.03}_{-0.04}$	$0.20^{+0.08}_{-0.09}$ / $0.18^{+0.07}_{-0.07}$ / $0.21^{+0.07}_{-0.09}$
Upper bound on primary spin magnitude a_1	0.74 / 0.65 / 0.62	0.74 / 0.76 / 0.72
Upper bound on secondary spin magnitude a_2	0.89 / 0.69 / 0.67	0.87 / 0.80 / 0.72
Lower bound on mass ratio q	0.68 / 0.69 / 0.70	0.48 / 0.40 / 0.55

We perform controlled tests by injecting synthetic GW signals into zero noise, reconstructing their parameters using `NRSur7dq2` templates as filters, and comparing the parameter recovery with both the true parameters as well as what other waveform models furnish. We perform a total of 48 such injections, which are summarized in Table I and use `NRSur7dq2HM` (full multi-modal surrogate), `NRSur7dq2` (dominant-mode only surrogate), IMRPhenomPv2 (a model for spin-precessing binaries with reduced spin degrees of freedom) and SEOBNRv4 (a model for spin-aligned binaries). We use IMRPhenomPv2 and SEOBNRv4 models in two configurations - first, where they are restricted to sample from the same prior parameter ranges that numerical surrogates are confined to, and second, where they can sample from a larger parameter range. The injected source parameters are varied as follows: mass ratio takes on values $\in \{1.2, 1.5\}$ while the total mass is fixed to $60M_\odot$; source spins are chosen from four strongly precessing configurations; source distance is varied over $\{500, 1000, 1500\}$Mpc, while source inclination is allowed two values - nearly face-on and nearly edge-on; and sky location angles are chosen uniformly over a 2-sphere. We find that both with and without the inclusion of $\ell > 2$ waveform multipoles, the `NRSur7dq2` templates are broadly consistent with the comparable IMRPhenomPv2/SEOBNRv4 in terms of parameter recovery. Some of the dominant parameters that describe any observed GW signal - sky location and chirp mass - are recovered equally well by all models. However, we find that the degeneracy in measuring source distance and inclination is largely reduced with the addition of $\ell = \{3, 4\}$ multipoles in `NRSur7dq2HM` and with it we can recover both of these parameters better than all other models, including `NRSur7dq2`. In addition, we also find that the total mass and the dominant spin contribution - the effective spin - are also better recovered using information from high-order modes. While past work applying higher-mode information from NR to GW parameter estimation could only use it to measure a subset of source parameters and relied on the interpolation of likelihood on an unstructured grid [23, 24, 30], the tests described above demonstrate the first *comprehensive* usage of higher-mode information from NR without additional approximations. Most of this improvement is moderate,

however, and we expect it to become more pronounced as the signals themselves have either a larger relative contribution from $l \neq 2$ modes, such as for binaries with higher mass ratios, or larger spin magnitudes. In the latter case, however, the sampling priors imposed on spins can alter their estimation appreciably [38] and must be carefully chosen. Our results provide motivation to extend the `NRSur7dq2` model to a larger range in binary mass ratios.

From these controlled tests, we establish the viability of using NR information directly in a traditional Bayesian parameter estimation framework for GW events. We next proceed to analyze the first BBH merger event ever to be recorded: GW150914. We analyze it with our NR surrogates - `NRSur7dq2HM` and `NRSur7dq2` - in addition to IMRPhenomPv2 and SEOBNRv4. The latter two (or their variants) were used in the original published analyses for this binary [37]. We find that with `NRSur7dq2HM` we place the source of this event to be at a luminosity distance of $\simeq$ 530Mpc, which is about 25% further away than what other models estimate (and previous LVC analysis of the event [37]). If we remove $\ell = \{3, 4\}$ modes and restrict the analysis to `NRSur7dq2`, we find the results agree with the originally estimated value, indicating clearly that this new information is extracted by the sub-dominant waveform multipoles in our filtering model. Simultaneously, `NRSur7dq2HM` also helps narrow down the allowed inclination configurations for the source to be either face-on or face-off with more confidence than earlier. Both of these improvements can be seen from Fig. 5. As would be consistent with a larger luminosity distance, the NR surrogates estimate the chirp mass of GW150914 to be marginally higher than the model estimate. Consistent with a larger redshift, `NRSur7dq2HM` estimates GW150914's mass in its source frame to be approximately 1% lower than other models. Finally, with fully nonlinear spin information implicitly contained within them, the surrogate models constrain the effective-spin of GW150914 better around $\chi_{\rm eff} \simeq 0$ than earlier estimated. Components of spin that are orthogonal to the orbit and that contribute to its precession are not constrained much better than phenomenological models, and this is as expected because for short signals there is simply not enough time for the binary to complete a few precession cycles. However, measurement of BH spins can be sensitive to the choice of sampling priors employed [30, 38]. We defer an investigation of their effect on spin inferences with `NRSur7dq2HM` to future work.

Finally, we move on to the second heavy BBH merger event: GW170104. This event differs from GW150914 in the sense that its measured posterior probability distribution (by approximate models) has some support outside the region of validity of `NRSur7dq2HM`, specifically for binaries with mass ratios $q > 2$. In practice, however, we find that this restriction does not bias the parameter recovery by `NRSur7dq2HM` in any noticeable manner. Similar to GW150914, `NRSur7dq2HM` constrains the luminosity distance to this event to be approximately 20% larger than what approximate models that include only the dominant $\ell = |m| = 2$ modes give [4]. The orientation of this source is constrained to be close to either face-on or face-off with edge-on configurations being more strongly eliminated. The estimation of mass combinations is consistent between `NRSur7dq2` and other models, although the former can only reconstruct an incomplete posterior distribution for mass ratio. Lastly, we note that the estimation of spin parameters by `NRSur7dq2HM` remains remarkably similar to what we get from IMRPhenomPv2 and SEOBNRv4, and is therefore consistent with them. The effective spin for the event is the only well-measured spin parameter. It is constrained to be small but negative by all models, while for all other spin combinations all models essentially recover the sampling prior as the posterior, with data adding little information.

From our results it is clear that there are certainly advantages of using numerical relativity surrogates for following up heavy binary black hole merger events. These include resolution of the luminosity distance - inclination angle degeneracy. This degeneracy often leads to a systematic bias in providing point estimates of distance, which then subsequently percolates to the calculations of astrophysical binary merger rates [64], estimation of Hubble's constant [65], etc. These applications could potentially benefit from `NRSur7dq2`-based follow-ups of GW events. It is also clear that additional information is typically recovered from a source that has a relatively larger fraction of its total information (or signal power) in non-quadrupolar waveform multipoles. This is increasingly the case at higher binary mass ratios, and therefore one of the lines of active research we are currently pursuing is to extend the domain of `NRSur7dq2` to model binary emitters with mass ratios $q > 2$. Finally, we also demonstrate that spin degrees of freedom are often better constrained with surrogates as they include unabbreviated representation of this degree of freedom. Our results are encouraging and we propose for `NRSur7dq2` and its follow-up models to used in standard GW event follow-up analyses in order to maximize the science output from GW detector data. Finally, we provide full posterior samples (as supplemental materials) from Bayesian parameter estimation of LIGO/Virgo data for GW150914 and GW170104, with `NRSur7dq2HM` and IMRPhenomPv2, to enable further analysis by the community.

ACKNOWLEDGMENTS

We gratefully acknowledge support for this research at Cornell from the Sherman Fairchild Foundation and NSF grant PHY-1606654; at Caltech from the Sherman Fairchild Foundation and NSF grant PHY-1404569; at CITA from NSERC of Canada, the Ontario Early Researcher Awards Program, the Canada Research Chairs Program, and the Canadian Institute for Advanced Re-

search; and at Princeton from NSF grant PHY-1305682 and the Simons Foundation. SEF was partially supported by NSF grant PHY-1806665. PK would like to thank FAPESP grant 2016/01343-7 for partial support of his visit to ICTP-SAIFR from February-March 2018 where part of this work was completed. This research has made use of data, software and/or web tools obtained from the LIGO Open Science Center (https://losc.ligo.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].
[2] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 241103 (2016), arXiv:1606.04855 [gr-qc].
[3] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. **X6**, 041015 (2016), arXiv:1606.04856 [gr-qc].
[4] B. P. Abbott *et al.* (VIRGO, LIGO Scientific), Phys. Rev. Lett. **118**, 221101 (2017), arXiv:1706.01812 [gr-qc].
[5] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. Lett. **119**, 141101 (2017), arXiv:1709.09660 [gr-qc].
[6] S. A. Usman *et al.*, Class. Quant. Grav. **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].
[7] K. Cannon *et al.*, Astrophys. J. **748**, 136 (2012), arXiv:1107.2665 [astro-ph.IM].
[8] S. Privitera, S. R. P. Mohapatra, P. Ajith, K. Cannon, N. Fotopoulos, M. A. Frei, C. Hanna, A. J. Weinstein, and J. T. Whelan, Phys. Rev. **D89**, 024003 (2014), arXiv:1310.5633 [gr-qc].
[9] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade, Phys. Rev. D **91**, 042003 (2015).
[10] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 221101 (2016), arXiv:1602.03841 [gr-qc].
[11] F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005), arXiv:gr-qc/0507014 [gr-qc].
[12] F. Pretorius and M. W. Choptuik, J. Comput. Phys. **218**, 246 (2006), gr-qc/0508110.
[13] F. Pretorius, Class. Quantum Grav. **22**, 425 (2005), gr-qc/0407110.
[14] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006), arXiv:gr-qc/0511048 [gr-qc].
[15] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Phys. Rev. Lett. **96**, 111102 (2006), arXiv:gr-qc/0511103 [gr-qc].
[16] J. G. Baker, M. Campanelli, F. Pretorius, and Y. Zlochower, Class. Quantum Grav. **24**, S25 (2007), gr-qc/0701016.
[17] A. Buonanno, Y. Pan, H. P. Pfeiffer, M. A. Scheel, L. T. Buchman, and L. E. Kidder, Phys. Rev. D **79**, 124028 (2009), arXiv:0902.0790 [gr-qc].
[18] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. H. S. Husa, M. Koppitz, D. Pollney, L. Rezzolla, L. Santamaría, A. M. Sintes, U. Sperhake, and J. Thornburg, Class. Quantum Grav. **24**, S689 (2007), arXiv:0704.3764 [gr-qc].
[19] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Phys. Rev. **D92**, 023002 (2015), arXiv:1502.04370 [gr-qc].
[20] M. Boyle, D. A. Brown, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, G. B. Cook, and S. A. Teukolsky, Phys. Rev. D **76**, 124038 (2007), arXiv:0710.0158 [gr-qc].
[21] P. Kumar, T. Chu, H. Fong, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, M. A. Scheel, and B. Szilágyi, Phys. Rev. **D93**, 104050 (2016), arXiv:1601.05396 [gr-qc].
[22] P. Kumar, K. Barkett, S. Bhagwat, N. Afshari, D. A. Brown, G. Lovelace, M. A. Scheel, and B. Szilgyi, Phys. Rev. D **92**, 102001 (2015), arXiv:1507.00103 [gr-qc].
[23] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. D **94**, 064035 (2016), arXiv:1606.01262 [gr-qc].
[24] J. Lange *et al.*, Phys. Rev. **D96**, 104041 (2017), arXiv:1705.09833 [gr-qc].
[25] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Phys. Rev. X **4**, 031006 (2014), arXiv:1308.3565 [gr-qc].
[26] M. Pürrer, Class. Quantum Grav. **31**, 195010 (2014), arXiv:1402.4146 [gr-qc].
[27] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Phys. Rev. Lett. **115**, 121102 (2015), arXiv:1502.07758 [gr-qc].
[28] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, Phys. Rev. **D95**, 104023 (2017), arXiv:1701.00550 [gr-qc].
[29] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilgyi, Phys. Rev. **D96**, 024058 (2017), arXiv:1705.07089 [gr-qc].
[30] J. Lange, R. O'Shaughnessy, and M. Rizzo, (2018), arXiv:1805.10457 [gr-qc].
[31] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, Phys. Rev. D **84**, 024046 (2011), arxiv:1012.2879.
[32] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015), arXiv:1408.1810 [gr-qc].
[33] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Prrer, X. Jimnez Forteza, and A. Boh, Phys. Rev. **D93**, 044007 (2016), arXiv:1508.07253 [gr-qc].
[34] L. Pekowsky, J. Healy, D. Shoemaker, and P. Laguna, "Impact of higher-order modes on the detection of binary black hole coalescences," (2012), arXiv:1210.1891 [gr-qc].
[35] J. Healy, P. Laguna, L. Pekowsky, and D. Shoemaker, (2013), 10.1103/PhysRevD.88.024034, arXiv:1302.6953

[gr-qc].
[36] J. Caldern Bustillo, S. Husa, A. M. Sintes, and M. Prrer, Phys. Rev. **D93**, 084019 (2016), arXiv:1511.02060 [gr-qc].
[37] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].
[38] K. Chatziioannou, G. Lovelace, M. Boyle, M. Giesler, D. A. Hemberger, R. Katebi, L. E. Kidder, H. P. Pfeiffer, M. A. Scheel, and B. Szilgyi, (2018), arXiv:1804.03704 [gr-qc].
[39] `http://www.black-holes.org/SpEC.html`.
[40] D. Gerosa, F. Hbert, and L. C. Stein, Phys. Rev. **D97**, 104049 (2018), arXiv:1802.04276 [gr-qc].
[41] C. Talbot, E. Thrane, P. D. Lasky, and F. Lin, (2018), arXiv:1807.00990 [astro-ph.HE].
[42] S. A. Smolyak, in *Dokl. Akad. Nauk SSSR*, Vol. 4 (1963) p. 123.
[43] H.-J. Bungartz and M. Griebel, Acta numerica **13**, 147 (2004).
[44] M. Boyle, R. Owen, and H. P. Pfeiffer, Phys. Rev. D **84**, 124011 (2011), arXiv:1110.2965 [gr-qc].
[45] R. O'Shaughnessy, B. Vaishnav, J. Healy, Z. Meeks, and D. Shoemaker, Phys. Rev. D **84**, 124002 (2011), arXiv:1109.5224.
[46] A. Buonanno and T. Damour, Phys. Rev. D **59**, 084006 (1999), arXiv:gr-qc/9811091 [gr-qc].
[47] L. Blanchet, Living Rev. Rel. **17**, 2 (2014), arXiv:1310.1528 [gr-qc].
[48] K. D. Kokkotas and B. G. Schmidt, Living Rev. Rel. **2** (1999), 2.
[49] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **95**, 044028 (2017), arXiv:1611.03703 [gr-qc].
[50] L. S. Collaboration, "LSC Algorithm Library software packages LAL, LALWRAPPER, and LALAPPS," .
[51] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **89**, 084006 (2013), arXiv:1307.6232 [gr-qc].
[52] S. Vitale, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, I. Mandel, B. Aylott, and J. Veitch, Phys. Rev. **D85**, 064034 (2012), arXiv:1111.3044 [gr-qc].
[53] T. B. Littenberg, M. Coughlin, B. Farr, and W. M. Farr, Phys. Rev. D **88**, 084044 (2013), arXiv:1307.8195 [astro-ph.IM].
[54] G. M. Harry (LIGO Scientific Collaboration), Class. Quantum Grav. **27**, 084006 (2010).
[55] J. Aasi *et al.* (LIGO Scientific Collaboration), Class. Quantum Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
[56] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, *Proceedings, 10th International LISA Symposium: Gainesville, Florida, USA, May 18-23, 2014*, J. Phys. Conf. Ser. **610**, 012021 (2015), arXiv:1410.4839 [gr-qc].
[57] J. Skilling, AIP Conference Proceedings **735**, 395 (2004), https://aip.scitation.org/doi/pdf/10.1063/1.1835238.
[58] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Class. Quantum Grav. **34**, 104002 (2017), arXiv:1611.07531 [gr-qc].
[59] T. D. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. **X6**, 041014 (2016), arXiv:1606.01210 [gr-qc].
[60] P. B. Graff, A. Buonanno, and B. S. Sathyaprakash, Phys. Rev. D **92**, 022002 (2015), arXiv:1504.04766 [gr-qc].
[61] L. London, S. Khan, E. Fauchon-Jones, C. Garca, M. Hannam, S. Husa, X. Jimnez-Forteza, C. Kalaghatgi, F. Ohme, and F. Pannarale, Phys. Rev. Lett. **120**, 161102 (2018), arXiv:1708.00404 [gr-qc].
[62] P. A. R. Ade *et al.* (Planck), Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
[63] J. Aasi *et al.* (the LIGO Scientific Collaboration, the Virgo Collaboration), Phys. Rev. D **88**, 062001 (2013), arXiv:1304.1775 [gr-qc].
[64] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Astrophys. J. **833**, L1 (2016), arXiv:1602.03842 [astro-ph.HE].
[65] A. Nishizawa, Phys. Rev. **D96**, 101303 (2017), arXiv:1612.06060 [astro-ph.CO].